

Europe still unsure of itself

The past few weeks have produced yet more evidence that European states are unsure where Europe as a whole is heading. They should make up their minds.

THE great excitement in Britain in the past two weeks about the fate of a small and unsuccessful helicopter company has been presented as part of the struggle for a United States of Europe (see *Nature* 319, 86; 1986). Mr Michael Heseltine, Secretary of State for Defence two weeks ago, walked out of the British Cabinet for all kinds of reasons, some of which may have been strictly personal, but ostensibly because he said his colleagues in the government were less than fair towards his scheme for letting control of the helicopter company pass to a European rather than to a US-Italian consortium. The general assumption seems to be that those who favour the European bid are true Europeans, and that the supporters of the Sikorsky-Fiat deal, by contrast, are lukewarm or even hostile to Europe. Everybody seems to have overlooked the truth, that the notion that Westland helicopters should be "saved" for Europe by a hastily constructed financial bid is a measure, rather, of the stark inadequacy of Europe's economic integration so far. For in a rational Europe, tears might be shed about the disappearance of a single helicopter manufacturer, but potential purchasers of these machines (of which there are few) would know that there are other companies to which they could turn for supplies.

Much the same applies to the announcement on Monday this (see p.251) week of the agreement between France and Britain that there should be a fixed link across the strip of water that divides them. Officials in Paris have made heavy weather, in the past few weeks, of their attempt to choose one from among the several competing schemes for forging such a link. Because civil servants are not renowned as engineers, it is not surprising that their technical appraisal has been tortuous and that its outcome commands very little confidence. The mere fact that the two governments have been drawn into the adjudication is, however, a confession of the incompleteness of the European programme. In a properly organized Europe, the question of whether private capital should be used to build a bridge, a tunnel or some hybrid of the two would no doubt have been the subject of a public administrative inquiry (since local interest in both countries will be affected), but the technical feasibility of the scheme would have been left to a decision by the markets in which the builders will have to raise the money.

The account on p.255 of the surplus of teachers in West Germany points from a different angle towards the conclusion that Europe is still an awkward compromise between the ideal and the politically possible. For a variety of reasons, West Germany seems to have continued training teachers, many of them specialists who would have taught in high schools, for longer than could be justified by the decline of the birth-rate that seems to have been general in Western Europe during the 1970s. But elsewhere within the European Communities, there are shortages of teachers. Even in Britain, where successive governments have acted vigorously to save public money by cutting back on teacher training and have created a climate in which recruitment to the teaching profession is difficult, there is now the prospect that the acute shortage of science teachers of the past few years will become chronic. Why do not out-of-work West German teachers move to Britain to fill these empty posts? Language difficulties are not a sufficient explanation. Only the

general sense that Europe is still an entity largely in name, which affects both individuals and institutions, can explain this immobility in the face of strictly parochial misfortune.

What is to be done? Last week, the European Parliament in Strasbourg gave its blessing to the moderate programme of reform of European institutions that has been negotiated over the past nine months between the governments of the European Communities. The effect will be to remove some of the impediments that now exist to the free movement of goods and services throughout Europe. But the reform is far less radical than some governments (that of France especially) had been crying for last year. It is far less than that required to make Europe an integrated economic unit — yet, even so, there is a chance that Denmark, which is committed to holding a referendum on the issue, will be found blocking this modest programme when it comes for decision in the Council of Ministers later in the year.

If these prevarications and inconsistencies persist, Europe's bluff that it is a continent will soon be called. When the Treaty of Rome that created the European Communities was negotiated in the early 1950s, everybody acknowledged that the new Europe would not spring into being overnight. Now, with more than a third of a century gone, the every-day spectacle of quarrels about the retention of customs posts (one of the matters glossed over in last week's Strasbourg vote) squares badly with proposals such as that of Mr Heseltine that European taxpayers should dig into their pockets to keep a helicopter manufacturer artificially in business for fear of undermining a purportedly integrated European defence industry. United Europe is now not much more than a collection of separate nations agreed to conceal their differences from each other beneath a cloak of even more dedicated extra-European chauvinism. That a united Europe would be an inward-looking Europe has always been a danger; the reality of a disunited xenophobic Europe is an even meaner outcome.

Mercifully, the other side of the coin is now wholly blank. Some European schemes have worked reasonably well. Participants in the Esprit programme for collaborative information technology, for example, say they have won benefits from joint support from Brussels that would not otherwise have arisen. They seem either not to know or not to care that they echo the public statements of participants in the early Euratom programmes of the 1950s, who were invariably well disposed towards unearned income. On this occasion, of course, things may turn out differently. Technological Europeans, in other words, should acknowledge that nothing much has yet been accomplished but that something may be. In other fields, much the same is true; even the freedom of people to move from one place to another only half restores *de jure* what was *de facto* commonplace two centuries ago (when "Europe" was twice as big).

The effective remedy therefore may be the calling of the European bluff. What can Europe do for European voters that national governments do not at present accomplish? The simplest answer is that intra-European trade has increased in the past 30 years, but this makes little sense because it came at a time when world trade as a whole was growing; since the faltering of world trade a decade ago, crises like that at Westland Helicopt-



ters have become two a penny. That is why Mr Heseltine's supporters should tell him that while his heart is in the right place, his head is in the clouds; that battling for a united Europe with Westland Helicopters as a slogan is fruitless. Mr Heseltine's local difficulty is that his supporters are not, like him, Europeans by instinct but rather xenophobes. He will not be the first to find himself caught out like that. □

No more numbers game

Last week's Soviet arms control proposals are the best yet, perhaps the best for forty years.

WASHINGTON is still forgivably hesitant and uncertain about last week's proposals on strategic arms control, which are breathtaking in their radicalism. Such a break with the recent past is bound to catch people off guard, which is as it should be. But it would be a great mistake to write off Mr Mikhail Gorbachev's suggestions for what should happen next at Geneva as yet another proposal which, when picked apart, will be found to contain little of substance, or be otherwise objectionable. For the novelty in what Mr Gorbachev had to say last week is his explicit acknowledgement that substantial controls on strategic weapons will not be reached quickly, and that it may take until the end of the century for the abolition of nuclear weapons to be feasible. Previous Soviet leaders have so often impeded rational discussion with impatient demands for "general and comprehensive disarmament" the day after tomorrow that the influence of a pragmatist should be generally welcomed.

The further merit of the Gorbachev plan, with several independent components such as the withdrawal of Soviet and US missiles from the European theatre, a reduction by a half of strategic arms and so on, is that the scheme is plainly not meant as a package but as a negotiating framework. If it is the will of the superpowers to reduce the numbers of strategic weapons they deploy against each other, for their mutual benefit and that of bystanders, it is clear that progress can be made only step by step, and that the first steps must be taken before even the character of later steps is known. The trick is to ensure that none of the intermediate steps will be obstacles to those certain to follow. The negotiators at Geneva are hardly in a position to take such a grand view, which is why there is a need for a response from none other than President Reagan himself.

There are two snags, of which the first is the obvious danger that the future ideal may undermine what is feasible now. It is clearly essential that the goal of abolishing nuclear weapons by the end of the century should not distract attention from tangible steps that could be taken now. The Geneva negotiators who reassembled last week could do worse than settle for a tidying up of the agreements to which they are already parties, SALT II and the Anti-Ballistic Missile Treaty in particular (see *Nature* 319, 164; 1986), which would seem far from an empty achievement if there is a firm promise of better things to come.

But does the world, or anybody, wish to see the back of nuclear weapons altogether? That is the second snag that everybody must face. The immediate reaction to the Gorbachev proposals from Mr Zbigniew Brzezinski, President Carter's National Security Advisor, that the scheme is a recipe to "make the world safe for conventional war", points to the nub of the difficulty. Nuclear weapons have become so much an aid to the preservation of the *status quo* that even the most farsighted statesmen will think twice before suggesting their abolition. Only the most idealistic or naive believe that the abolition of nuclear weapons could be finally accomplished without agreement on the uses to which conventional forces would be put. But issues like that even touch the degree to which national sovereignty is an autonomous force in international relations. Perhaps it is not too soon to begin asking questions like these, but their difficulty should warn even the optimists that Mr Gorbachev's timetable, worthy though it is, will be much delayed. □

Scribes in space

The US government plans to put a journalist in space later in the year. Is it wise?

SOME years ago, US journalism was enlivened by a stocky character with a Brooklyn accent whose career had been transformed by a single assignment, that of being the first journalist to witness in that capacity the explosion of a nuclear weapon at a Pacific atoll. William L. Lawrence by birth, he was known to everybody as "Atomic Bill", and required by the expectations of his companions and editors to fill the role, whatever his inclinations may have been. For decades afterwards, no nuclear reactor could go critical, no diplomatic conference on the future of nuclear energy could take place, without his presence. Only when he retired from the *New York Times* to a Mediterranean island was Atomic Bill able to follow his chief interest, pre-nucleic acid biochemistry. But now, with characteristic indifference to the lessons of history, the National Aeronautics and Space Administration (NASA) is about to wish a similar fate on another unsuspecting journalist, the first ever to be sent up by shuttle.

By the closing date for applications last Thursday, the US Association of Schools of Journalism and Mass Communication, rising to the only occasion in its history likely to be remembered generally, was said to be swimming in a large proportion of the 4,000 application forms requested by aspiring Orbit Micks (or Megs, as the case may be), many of them turned in, in the best traditions, right on deadline. It is hoped that not too many readers will cancel their subscriptions because none of the members of *Nature's* staff who satisfies the condition of being a US citizen will also meet the requirement of having been in journalism for five years. But there are more onerous conditions.

The worst is that the unlucky winner of the competition will have to work as a one-person "pool", a kind of communal resource to be exploited by all his or her natural competitors, who will be free to demand answers to whatever questions come into their heads. ("Say, that nausea, is it worse when you move about?") With NASA in charge of all communications, the pool reporter's traditional ingenuity in keeping the best scraps of information for his own employers will be fully tested. NASA's demand that the reporter should "not intrude on the private lives" of his companions "without their express permission" will not easily be satisfied. It will be irksome that personal equipment that cannot be "flight-qualified" will have to be replaced by what NASA thinks best. The summer in Houston is not everybody's idea of bliss.

Who will be chosen? This is where the fun will start. The medical examination seems not much of a hurdle. The need for a security clearance may be a more serious obstacle even though NASA promises that recruitment will not be based on "political associations or beliefs". But the real fun starts with the two "essay questions" that applicants must answer. One whole sheet of blank paper awaits the response to "How will you attempt to fulfill" the purpose of communicating the "unique experience of space travel"? Maybe the selection panels, representing as they do the schools of journalism and mass communication, have ways of telling when people who fill this sheet with phrases like "mankind liberated from tedious Earth" will lapse, in the event, into phrases like "I can see Paris, France!" But the trick question is the second, which asks that the chosen journalist should speculate, "looking 10 or 20 years into the future", about the value to journalism of the facility "to report regularly from space". It is as if a restaurant were to offer free dinners to those with interesting opinions on the value of dining out. Cruelly, NASA retains the right to publish these essay questions, ensuring that it can ruin not just the career of its equivalent of Atomic Bill, but all those who answer this question with a straight face. □

Archaeologists

World congress quits Southampton for Mainz

THE University of Mainz (West Germany), rather than Southampton (England), will host the 11th World Archaeology Congress in September 1986. The executive committee of the International Union of Prehistoric Sciences (IUPPS), at an extraordinary meeting in Paris last week, took the unprecedented step of "withdrawing recognition" of the congress from the University of Southampton.

Professor John Evans, chairman of the executive committee and president of IUPPS, resigned his position during the Paris meeting, but is to stay on as an ordinary member of the executive. Professor Peter Ucko, organizer of the Southampton congress, ceased to be a member of the executive after the decision to remove the congress from Southampton and left the Paris meeting before it ended.

IUPPS's decision resolves the bitter dispute caused by the banning of South African delegates, including scientists of other nationalities who work in South Africa, from the Southampton meeting (see *Nature* 317, 754; 1985). The ban resulted from pressure by Southampton's city council and university teachers' and students' trades unions, although Ucko argued strongly against it.

More than 200 participants from the United States, as well as the delegations from France, West Germany, The Netherlands, Belgium, Switzerland and Spain, would have boycotted the meeting

Do you think that if we dug deeply enough in South Africa, we'd find any trace of civilization?



had the ban been upheld, whereas many African delegates would not have attended had the ban not been introduced. The IUPPS executive took its action because the organizing committee of the Southampton meeting had departed from IUPPS's statutes and longstanding policy to keep all its activities open to all "bona fide scholars". For more than 50 years, the 10 congresses sponsored by IUPPS have been organized on this basis and scholars have been admitted irrespective of race, nationality, religion or political convictions.

The new president of IUPPS and chairman of the executive committee is Profes-

sor K. Böhner of Mainz, and Dr K. Weidemann is the organizing secretary. Notice of the decision to move the congress is to be sent to the members of the permanent council of IUPPS for approval by postal vote.

Professor Phillip Tobias, a member of the council, said in Paris this week that the action "constitutes a landmark decision in the history of the struggle for the freedom of science and for the free circulation of scientists. . . . The IUPPS resolutions constitute a striking reaffirmation of the need to keep scientific intercourse and the access to knowledge free from political, racial or geographical limitations."

Maxine Clarke

Searle's UK research lab closes

G.D. SEARLE, the US drug company, is to end its pre-clinical research programme in the United Kingdom by closing its High Wycombe laboratories with the loss of 300 jobs. The decision was made largely to avoid a "duplication of effort" in company laboratories in other countries and is part of a worldwide reorganization of the company in the wake of Searle's recent merger with the giant US chemical company Monsanto. At that time, Searle said that the research enterprises of the two companies would reinforce each other.

Much of Searle's UK research effort will be absorbed by Monsanto's directly comparable biotechnology research institute in St Louis, Missouri; the rest will go to Searle laboratories in Belgium and France. Staff will be offered transfers on an "individual basis".

Searle's decision to stop its UK programme stems partly from its disillusionment with the British government's recent regulations designed to save money on National Health Service drugs prescriptions. British-owned companies get a better price for drugs than foreign companies. But a company spokesman stressed last week that Searle will spend more this year on research and development worldwide than ever before; \$160-170 million in 1986 compared with \$130-140 million in 1985. And individual projects in universities, for example a £1 million project at the University of Oxford, will continue. Searle is now looking for a new site in High Wycombe for its business operations — the present laboratories are considered "unsuitable" for that purpose. Maxine Clarke

Channel tunnel

Rail plan gets go-ahead

AFTER nearly two centuries of talking, Britain and France are to be linked by a tunnel under the English Channel (*La Manche*, the "sleeve", to the French). A historic moment? Not to judge by the French media, which have taken very little interest in the prospect of getting from Paris to London in three and a half hours rather than five.

This is understandable. Through the tunnel, Britain sees the whole of Europe: France sees only Britain. And what do the French think of Britain? In a recent poll by a French science magazine, readers were asked with which country they would like to see French industry and scientists cooperating in high technology. Nearly 30 per cent said West Germany. Only 1 (yes, 1) per cent said Britain.

But, notwithstanding popular French indifference, Prime Minister Margaret Thatcher and President François Mitterrand announced on Monday that construction can now begin of a twin-bore rail tunnel, though which trains will carry freight direct and cars piggy-back. The rail tunnel should be open by 1993. This looks like defeat for the "road" lobby, Mrs Thatcher included, although the winning consortium, the Channel Tunnel Group (CTG), will be asked to prepare plans for a twin road tunnel to be built within fifteen years — if it appears economic.

The decision has not pleased those placed either side of the tunnel. The mayor of Calais, the principal French ferry port, has said that within a few years of the opening of the tunnel Calais would become "a ghost town". That might be true, and the same for Dover, if the rail tunnel provides transit at costs much lower than the ferries, as predicted by CTG's chairman, Sir Nicholas Henderson. A spokesman for the now-defunct "Euroroute" bridge scheme, however, claims that drilling problems and heating and ventilation of a tunnel will force up costs far beyond those now predicted.

A more optimistic scenario, however, has it that the "chunnel" will indeed reduce transit costs, and thus encourage new industries to set up near the tunnel entrances. However, the zone south of the tunnel (in north-east France) may well prove more attractive than that north of it (in Kent and Sussex), as the former will also provide a crossroads for traffic moving east and west across Europe.

French rail interests are also keen on the prospect of running the advanced French high-speed train, TGV (*Train Grand Vitesse*) to London. In many ways, it seems, Monday's decision looks like a French success.

Robert Walgate

Japanese science budget

Some gains, some losses

Tokyo

BESIDES some big funding increases for a handful of key projects, there is little good news for scientists in the 1986 budgets of two of Japan's major science ministries, the Science and Technology Agency (STA) and the Ministry of International Trade and Industry (MITI).

This year, as for the past few years, the government has done everything it can to cut the size of the budget. What makes this necessary is the enormous fiscal deficit, proportionately just as big as that of the United States. And despite Japan's booming economy, the alternative of raising taxes is no more popular in Japan than in the United States. All ministries are under pressure to make cutbacks and the fact that there have been small increases for science reflects its high priority. The small gains in research and development budgets are in fact scarcely enough to cover inflation. But that has not stopped the launching of new projects and the concentration of funds in key areas.

Within STA's budget, the biggest leap of all (a 750 per cent increase) is seen for the construction of a bathyscaphe capable of diving to perhaps 6,500 m. The plan to build such a vessel — at present only France and the United States have a bathyscaphe capable of going that deep — has been around for many years. With the successful completion in 1981 of Shinkai 2000, which can reach 2,000 m, it has become a logical step. The new vessel will allow direct observation of the deep troughs off Japan where many earthquakes originate.

Another major budget increase is for planning of a module to be attached to the US space station. The budget is still not

large in real terms, however, and if the project is delayed by the United States it may not prove too unwelcome to the government. If the module is built, a major use will be the manufacture of new materials in a gravity-free environment. MITI has launched a new project on this subject with an initial budget of 200 million yen. Interest from industry is strong enough that it may find it has a commercial competitor one day: the Taisei construction company is thinking of building a gravity-free environment on Earth. The trick is to carry out the manufacturing process in a capsule in free-fall down a deep mine, or from the top of a tall building.

Another area of space research where Japan is aiming to acquire expertise is the construction of the H-II rocket. Although Japan's large launch vehicles have so far been mainly built under licence, STA's H-II is seen as competing with Ariane and the space shuttle for commercial satellite launches towards the end of the century. This year its budget has leapt by nearly 300 per cent.

Another growth area for STA is in a project to build a heavy-ion accelerator to treat cancer. Funds available for basic research grants are also up, although not by such a huge percentage, and some new funds are available for "frontier research". These confirm STA's role as a supporter of small-scale basic research — a position traditionally held by the Ministry of Education, Culture and Science — as well as a supporter of "big science", space, nuclear energy and the like.

MITI's budget shows continuing support for its large-scale projects — the Sunshine project to develop new energy sources, the Moonlight project to cut energy waste, particularly in large-scale manufacturing, and the set of industrial technology projects. A new theme to be added to the "Basic technologies for future industry project" is that for bio-electronics. Although application is a long way in the future, there has been considerable interest from several large companies, including NEC. Interest stems from twin beliefs that computer manufacturers can learn from the structure of the nervous system, which is why MITI's industrial research agency supports physiology research, and that biological systems might be worth studying as the ultimate in miniaturization.

One project that is not growing as originally planned is the Fifth Generation Computer Project. The belief seems to be that the project's limits are not being set by financial resources but by human ones — there is not enough expertise available to justify further massive spending at present.

Alun Anderson

Space research

Texas A & M fly space-grant kite

Washington

TEXAS A & M University, in an attempt to fill what it sees as a vacuum, is pushing for the creation of a space-grant programme, modelled on the concept of the land-grant and sea-grant programmes. If Texas A & M has its way, it will add space-grant to its land-grant and sea-grant designations.

The sea-grant programme, started in the late 1960s, is a cooperative effort between universities, industry and government to provide financial support for research and public information. The space-grant programme would extend this to include space-based research, as well as research performed on Earth to support long-term projects in space.

Frank Vandiver, president of Texas A & M, sees the space-grant programme as "a good way for government to maximize research dollars". Federal dollars should encourage participation by private industry in university projects, and the financial benefits of the collaboration should be shared by all. That, at least, is the idea.

The land-grant concept has had a profound effect on US higher education. In 1862, the Morrill Act provided federal lands to the states as an endowment for colleges teaching the mechanical and agricultural sciences. The emphasis on applied sciences was reinforced by later legislation providing for research activities and extension services for the dissemination of newly gained knowledge. The triad of teaching, research and public education forms the basis for most modern universities. The original land endowments now amount to less than \$3 million in university support, but the research and extension services are supported at levels reckoned in thousands of millions of dollars.

Whether space grant is a truly practical idea is unclear. The sea-grant programme's annual appropriation, this year \$37 million, is routinely eliminated from the budget by the Reagan administration, only to be restored by congressional supporters such as Senator Lowell Weicker (Republican, Connecticut). A space-grant programme would be yet another financial burden. Vandiver insists that the initial burden would be more than paid off by applications of sponsored research. He has persuaded Senator Lloyd Bensten (Democrat, Texas) to introduce legislation establishing a space-grant programme, although the Gramm-Rudman Act (see p.254) makes new spending programmes unattractive.

Some see the space-grant idea as a ploy to increase visibility in an era of tight budgets. But Vandiver insists "it is no flight of fancy with me".

Joseph Palca

Japan's science budget

	Yen (thousand million)	Percentage increase or decrease
Science and Technology Agency		
Research and development budget	427.8	+1.6
Special promotion funds	7.9	+8.2
Space	94.5	+3.3
Nuclear Energy	277.7	+4.2
Ocean Research	7.4	+7.3
ERATO	2.8	+4.6
Ministry of International Trade and Industry		
Research and development budget	197.0	+4.4
Agency of Industrial Science and Technology	113.4	+0.2
Basic technologies for future industries	6.5	+0.1
Large-scale industry	15.2	+3.4
Sunshine project	43.0	-1.9
Moonlight project	12.3	+10.8
Fifth-generation computer project	4.5	-6.3

Television

Defence of the realm

If further proof were needed that the British establishment is very cool to the idea of Europe, it came last week in the House of Lords report on international television. In answer to the European Commission's proposal to remove some of the restrictions on the flow of television across international barriers, the Select Committee on the European Communities responded with a disdain that can only keep the barriers up.

Not that the lords object to the objectives behind the European Economic Community's green paper of May 1984: *Television Without Frontiers*. Those are unimpeachable: to create a common market in television programmes and to create a sense of European identity. Nor would the subcommittee on energy and technology which took evidence for the report and included Lord Swann (former chairman of the British Broadcasting Corporation) and Lord Zuckerman wish to be seen to oppose the principle of free flow of information, still less to be trying to stop the advance of technology. The report recognizes the inevitable: broadband cable networks and direct-to-home satellite broadcasting will eventually blanket Europe. But the report's conclusions boil down to delaying tactics — not now, not here, not that way.

In fact, the Commission's own proposals to which the lords were objecting are quite modest: that members of the European Community countries come to an agreement on some common advertising standards (to remove such absurdities and barriers to television exchange as the ban on petfood advertising in Italy and on contact lens commercials in Britain). They would also impose a quite relaxed limit on the total amount of programming time (20 per cent) that might, within the Community, be devoted to advertising. They also proposed a possible way through the copyright thicket. As copyright agreements are bound by national borders, it is nightmarish for writers, producers, musicians to try to negotiate the rights for their programmes to be shown in another country or, in the case of Britain, on the mainland of Europe at all. US experience suggests that the new approach to the problem can only come through blanket licensing arrangements; the days of negotiating for each individual transaction disappeared with the arrival of the audio and video tape.

The lords committee, however, firmly rejected every proposal. Rejected in turn were: restricting television advertising time (that could inhibit new kinds of advertising channels); adopting a blanket copyright procedure (that would deprive copyright holders of their freedom). The

lords even rejected as unacceptable or unworkable two things the green paper did not propose: a total ban on alcohol advertising and a statutory right to reply for aggrieved parties. In sum, they found the Community the wrong body to regulate and said, contradictorily, that its proposed regulations were both premature and unnecessary as "viewers' freedom to choose will inevitably take the place of regulation".

The committee's strongest point was to argue that the Community is too small (even with 12 members) to adopt such regulations. If the new television environment has satellite signals disregarding all national boundaries, certainly wafting all across Europe, any code that ignores Austria, Norway and Switzerland will not mean very much.

In its confident nationalism, the lords' report sounds remarkably similar to the view of the BBC, the Independent Broadcasting Authority and the European

Broadcasting Union. It has fallen victim to the besetting European sin of thinking that in really vital areas (telecommunications procurement is another) national is best — particularly in broadcasting.

But the report, with the notable exception of its evidence, (one of the best summaries available of the hard-to-get information on the status of cable and satellite television in Europe) smacks of Little Englander thinking. The new standards would be minimum, not mandatory. If Britain wanted to restrict its own broadcast television to seven minutes an hour, it could still do so.

British broadcasting is probably the best in the world and regulation of the domestic product can keep it that way. But outsiders' television is crowding in. The Community's proposals are a step toward accepting that fact, and the even more uncomfortable one that if Europe is to become a real community, television programmes should be as free to travel as are workers.

Brenda Maddox

Television Without Frontiers (HL 43). House of Lords Select Committee on the European Communities: 4th Report 1985 – 86 HMSO, £12.

UK public spending

No pre-election sweetener

HOPES that the restraints on British research and education spending might be relaxed with the approach of the next general election (due no later than mid-1988) were dashed last week, with the publication of the government's estimates for public spending in the next three years (to April 1989).

The cash value of the science budget, mostly spent by the research councils, estimated at £585 million in the current year and due to rise by £15 million in real terms in the next year, will reach a total of £650 million in 1988–89, an increase that will be eaten away by an inflation rate of 3.5 per cent a year. The cash value of the government subvention of the universities over the same period similarly increases by just 10 per cent.

In spite of the pressures of the past few years, the British system of higher education is not yet actually shrinking, but has stabilized at a total of 484,000 students. Within this total, a decline in university student numbers of roughly 5 per cent (to a total of 256,000) has been offset by the steady growth of the polytechnic system. The government's white paper says that the "participation rate" (the proportion of the annual age group entering higher education) has risen to 14 per cent and will continue to increase for the remainder of the decade.

On research, the government says that its aim is "to maintain and enhance the strength and quality of the science base". The document says the government be-

lieves there is further scope for "greater concentration and selectivity", with more emphasis on value for money. The document includes a number of examples of British inventions that have been commercialized successfully, including the synthetic pyrethroids and the cephalosporins. The details of how funds will be distributed among the research councils will be settled, year by year, by the Advisory Board for the Research Councils.

Meanwhile, it seems plain that the squeeze on research will continue in those departments of state that have it within their control. Last year's decision by the Ministry of Agriculture, Fisheries and Food (MAFF) to cut back on research is the chief reason why agricultural research has been forced to continue cutting back on the work at its institutes; last week's document makes it plain that there will be no reprieve, for total spending is to be reduced to £90 million three years from now from the present £95 million, inflation notwithstanding.

Even defence research, the British government's chief outlet for research money in the past, will not be immune. Last week's statement shows that defence research and development spending will have fallen from 13.4 per cent of the total in 1979–80 to 12.8 per cent in the current year (when total UK defence spending is estimated at £18,222 million). □

The Government's Expenditure Plans 1986–87 to 1988–89 (House of Commons 9702 – I & II, HMSO, Vol. I £4.80, Vol. II £20.)

Gramm–Rudman Act

Automatic cuts threatened

Washington

WASHINGTON bureaucrats are slowly waking up to the uncomfortable realities of the Gramm–Rudman deficit reduction act approved by Congress at the end of last year. A comparatively modest 4.3 per cent across-the-board cut in non-defence spending for the current fiscal year will, barring some extraordinary legislative contortions, take effect on 1 March, together with a 4.9 per cent cut in defence spending, the saving will be \$11,700 million. But that is only a foretaste of the act's potential to wreak havoc. In order to avoid much larger cuts in future, totalling more than \$60,000 million in 1987, Congress is steeling itself for some uncharacteristic self-discipline.

The Gramm–Rudman Act seeks to eliminate the federal budget deficit, which was to amount to \$220,000 million this year, in five equal annual steps. If the President and Congress can agree each year on a budget that meets that year's target, all well and good. If not, the act mandates "sequestrations" to bring the deficit into line with the target. Defence spending has to carry half of the cuts, and there are exemptions for social security and welfare programmes.

The legislation has fundamentally changed the nature of the political game in Washington. Before, interest groups within the government could try to use their influence to secure their desired level of funding without much fear of adverse consequences. Now, they are having to learn that failure to agree would take the matter out of their hands. The first taste of serious deficit cutting will come next month when President Reagan delivers to Congress his proposed budget for 1987, which will attempt to meet Gramm–Rudman's \$144,000 million deficit target and so avoid automatic cuts.

Dr John McTague, acting director of the White House's Office of Science and Technology Policy, was nevertheless taking an optimistic view last week of the prospects for research. McTague pointed out that this year's 4.3 per cent "cut" was on a base already 7 per cent above last year's. As for future years, McTague said basic research had always been one of President Reagan's top priorities and observed that "we have a remarkably consistent President". McTague justified the President's decision to exempt the Strategic Defense Initiative (SDI) from cuts this year on the grounds that SDI "serves a different purpose" from other research.

National Science Foundation (NSF) officials are similarly cheerful, with good reason. The President's 1987 budget request includes a healthy 9 per cent increase for the foundation over last year's

total of \$1,524 million, which is likely to cause envy among less fortunate agencies. NSF is praying that Congress will manage to pass a budget that avoids the need for Gramm–Rudman's across-the-board cuts next year, which could at a stroke change a 9 per cent increase to an estimated 10 per cent decrease. Erich Bloch, NSF's director, last week termed the automatic cuts provision an "unintelligent" approach to budget planning.

The National Aeronautics and Space Administration (NASA) is in a much less comfortable position. This year's 4.3 per cent cut could lead to lay-offs and even the postponement of shuttle missions unless Congress passes legislation allowing NASA to transfer funds between its four main accounts, research, flight operations, salaries and facilities.

Looking to the future, the brunt of the savings being sought will fall on the space station, leading to a delay of 2–3 years for the project. NASA officials do not deny reports that despite an original 1987 budget request of \$580 million, the Office of Management and Budget is unwilling to countenance anything above \$100 million.

The National Institutes of Health (NIH) are as yet undecided about how they will cope with this year's cut. But it seems likely that they will try to avoid cutting back on numbers of grants, and instead reduce the size of the cheques.

Each of the various institutes will have to reduce its budget by 4.3 per cent in total, but directors will have discretion to redistribute funds within their own budgets to minimize disruption. NIH officials seem not to have thought seriously about what they would do if faced with large cutbacks next year; NIH usually end up with far more than the administration requests anyway.

Non-nuclear research at the Department of Energy (DoE) will be hit hard, with cuts in 1987 of more than 50 per cent from this year's total of \$700 million.

The Department of Agriculture says it will not have much difficulty in achieving 4.3 per cent cuts in the Agricultural Research Service this year, and that savings from vacant posts will probably account for many of them. But if across-the-board cuts are required in 1987 the results will be "too horrible to contemplate".

Many officials and congressional staff admit privately to doubts about whether Gramm–Rudman can survive for long: the savings sought are unprecedented in recent history. But while it may at some point be struck down on constitutional grounds, or simply eroded away by exemptions until it becomes meaningless, in the short term its effects will be real enough.

Tim Beardsley

Genetic engineering

Frost damage tests blocked

Washington

A NEW set of regulatory hurdles is impeding plans by Applied Genetic Sciences (AGS) of Oakland, California, to release bacteria, genetically altered to protect strawberry plants from frost damage, into the environment.

The problem now lies with the local government in Monterey County, California, where AGS plans to conduct its field trial. Sam Karas, chairman of the Monterey County Board of Supervisors, says residents are unhappy that the first trial is being conducted in a densely populated region. Local citizens are also upset at having no say in the matter. The board is seeking ways to block the trial until local concerns can be dealt with.

In addition, the Monterey Bay Unified Air Pollution Control District wants to see a detailed environmental impact statement before allowing the test to proceed, and has threatened legal action to block the trial if the information is not forthcoming.

The local furore has taken AGS by surprise. Despite the difficulty in obtaining federal approval, Douglas Sarojak, AGS director of product development and marketing, maintains that his company "did not realize the sensitivity to the issue". He adds, "We're scientists, not public relations experts."

AGS won an experimental use permit for its new strain of *Pseudomonas syringae* from the Environmental Protection Agency (EPA) on 14 November after a long struggle (see *Nature* 318, 96; 1985). The genetically altered bacteria lack a segment of the genome coding for an ice-nucleation protein responsible for frost damage to plants. The new strain will protect plants for competing with wild-type bacteria for positions on leaves.

The local controversy has taken on an international flavour. Karas reports that members of the board of supervisors have received telegrams from members of the Green Party in the West German Parliament as well as from members of the European Parliament urging them to oppose the tests. The board plans to hold a public hearing on 27 January to give all sides a chance to air their views.

Even if AGS receives permission from the local authorities, it may still not be able to proceed. The Foundation on Economic Trends in Washington led by Jeremy Rifkin has a suit pending in federal court seeking to overrule EPA approval of the experimental use permit for *P. syringae*, and Rifkin has gone to court to seek a temporary restraining order to prevent the field trial from starting before the lawsuit is settled.

Joseph Palca

Soviet Union

Scientists facing redundancy

MOST Soviet scientists will henceforth be paid in accordance with their job and performance. A recent joint decision of the Central Committee of the Communist Party of the Soviet Union, the Council of Ministers and the All-Union Central Council of Trade Unions approved new procedures and pay scales that will apply to more than half of all scientists and over 90 per cent of designers and technicians.

Although the new scheme has been presented as part of the Gorbachev package of reforms aimed at streamlining the economy, criticism in the Soviet media of the old system goes back a number of years. Under a set-up in which jobs and salaries depended strictly on academic qualifications and length of service, many scientists in the 35–45 age group, holding the degree of Candidate of Science (roughly equivalent to PhD), were actively or passively discouraged from improving their qualifications by taking a doctorate of sciences. In many cases, acquiring a doctorate would mean seeking work in another town, with consequent disruption to spouse's career and children's education. Scientists who wished, for their own personal satisfaction, to take a higher degree, and who were in some cases willing to continue temporarily in their old jobs and at their old salaries, were told that this would be detrimental to work discipline.

Under the new system, a degree or engineering diploma will be merely one of several criteria determining job and pay scales. From now on a scientist's pay will depend on the effectiveness of his or her research, its scientific and economic importance, and what Valentin Kharin, a senior official of the State Committee for Labour and Social Problems, described in an interview with the Moscow weekly *Literaturnaya Gazeta* as "personal impact" in getting concepts implemented in the economy. (The gap between research and implementation is a longstanding problem to the Soviet planners.) In particular, technicians and engineers without an academic degree seem likely to benefit. According to Mr Kharin, the new scheme will allow them to earn up to 350 rubles monthly, instead of the present maximum of 190 (the national average is around 170 rubles a month).

The new scheme makes likely the redundancy of up to 10 per cent of the present scientific work-force, with, Mr Kharin admits, some difficulty in finding new jobs for those who are dismissed. People who do not work, or who work badly, are liable to find themselves out of a scientific post, regardless of their length of service, he said. This is a remarkable development in Soviet terms. Unemployment was officially abolished in the Soviet

Union more than 50 years ago, and people without work are liable to find themselves facing prosecution for "parasitism" — being without visible means of support.

Compulsory retirement of senior scientists may be one way of weeding out the superfluous work force. Already, in Uzbekistan, the President of the local Academy of Sciences, Dr Abid S. Sadykov, has been eased out of his post. According to the Uzbek republic daily *Pravda Vostoka* (22 November 1985), by continuing in office for several years after

health and age (he is 72) made him unfit for duty, Sadykov had, by default, allowed corruption, nepotism and general slackness to creep into his academy. *Pravda Vostoka* noted disapprovingly that the average age of Uzbek academicians (75) is far too high and that more room should be made for younger talent. If this trend spreads, it could have important consequences for Soviet participation in world science. Attendance at foreign conferences has until now been viewed by the Soviets principally as a reward for long and faithful service. The Gorbachev efficiency drive could therefore produce a more flexible approach to foreign travel for younger scientists.

Vera Rich

West German schools

Teachers swell dole queues

Hamburg

UNEMPLOYMENT among West Germany's schoolteachers has become a scandal. The number has now reached 100,000, and is continuing to rise. The social cost is very large, given that a qualified teacher will have spent an average of six years at university, as well as perhaps two years at school practice (Referendariat).

Because of the waiting period of a year to eighteen months between university and Referendariat, the average age of qualifying teachers is as high as 30. Yet teachers are almost completely dependent on the state for jobs, in the person of the Kulturministers (the ministers for culture, education and church affairs) of the *Länder*. Fewer than one per cent of all teaching positions in West Germany are in private schools.

Several attempts are being made to solve the problems, but they are too late to help most teachers. Yet the present situation could have been foreseen many years ago from the crudest statistics. The Kulturministers had themselves published several documents giving calculations and predictions for the trends in supply and need of teachers, but nothing was done to prevent catastrophe. Until 1980, every newly qualifying teacher was assured of a post almost automatically. As civil servants, West German teachers are promised lifelong tenure.

The overproduction of teachers resulted from a campaign during the early 1970s when a shortage of schoolteachers of every type was emphasized. Too many students elected to become teachers, even when it became obvious that the shortages no longer existed. Even now, there are still too few teachers in some subjects, but the *Länder* do not want to spend more money in schools at which the numbers of pupils are decreasing sharply.

There have been protests from the teachers' professional organizations, chiefly the Gewerkschaft Erziehung und

Wissenschaft (Union for Education and Sciences) and the Deutscher Philologenverband (German Philologists' Society), but they represent only the employed teachers and there is little solidarity.

The ministers meet regularly, but they usually postpone discussion of this issue to their next meeting. They have not managed to answer even the most pressing question: how can the reduction of working hours — agreed for all civil servants during the union negotiations in 1984 — be applied to teachers? The absurd situation exists where one fewer teaching hour each week for all 590,000 teachers in post would provide jobs for 22,400 of their unemployed colleagues; education experts in both the Social Democratic and the Christian Democratic parties recommend such a reduction, but the civil servants have so far been successful in their opposition. They themselves are asking for more jobs to be created before they will agree to any kind of compromise.

The ministers, however, know that something must be done, even at this late stage. In Bavaria, all teachers over the age of 55 are to have their working week reduced by two hours. Nordrhein-Westfalia has given some applicants jobs for three-quarters of the time, but only on a three-year basis. Many unemployed teachers are, however, unwilling to give up jobs as taxi drivers, innkeepers, clerks or sales representatives only to face even bigger problems in three years' time.

At present, almost 5 per cent of West Germany's unemployed are teachers, but the proportion is growing dramatically. The vice-president of Munich University, Otto Speck, predicts an increase of 72,000 in 1986. The schools are facing "Vergreisung" (a growing percentage of older teachers), but the finance ministers of the *Länder* continue to argue that "teachers cannot be a privileged group among the unemployed".

Jürgen Neffe

West German universities

SIR—Your article on West German universities (*Nature* 316, 96; 1985) only reflects part of the actual situation.

The concept of "shifting balance within the university committees to a wider group" was indeed an integral part of the university reform that was rushed through some *Länder* parliaments in the early 1970s, without consulting academic staff. The anti-industry attitude then prevailing also led to the institutionalization of rigorous limitations on *Drittmittelforschung*, that is, industry-supported research. The idea was to reduce if not abolish access of individual staff members to research funds, and to ensure control of all departmental activities by "democratic" committees that were usually dominated by extreme left-wing activists.

The new legislation required all decisions concerning such divergent matters as professional appointments, finance, extension of contracts of laboratory assistants or undergraduate courses to be arrived at by majority vote in "democratic" committees. The resulting meetings occupied a considerable part of the staff's time — up to 90 hours per month was not unusual. Plans to abolish individual examinations as relics of capitalist exploitation and to accept examination papers produced by teams of comrades, so that the strong could support the weak, have been seriously discussed in this context.

This led to a decrease in research activities and further weakened the potential of German universities which had suffered from the exodus during the early years of Nazism, from twelve years of dictatorial rule and persecution and from the war. It is no accident that research papers from German universities are only rarely found in the pages of *Nature*, while the independent Max-Planck Institutes (MPIs) are now the main source of internationally recognized research in Germany. The same applies to the job situation — openings advertised in *Nature* are usually in MPIs and not in universities.

You refer to professorships that "were distributed too readily during the years of expansion"; the years of expansion were the 1960s, when common sense still prevailed; ready distribution of professorships took place under the new university regulations when pressure from the non-professional majorities on committees resulted, for instance, in the University of Hamburg elevating 400 lecturers to professors (not on merit!) by a single decree.

This is why prospects for promotion in universities are now discouraging and why the scientific standard of professors in "reformed" German universities does not compare favourably with the United Kingdom.

Let me conclude by saying that my friends in British universities are invariably horrified when they learn about the new German "system", which, in recent years, has also been adopted in my native Austria. A head of department is now a mere figurehead who has to maintain the goodwill of the powerful departmental committee by populist measures or by bowing to the majority of the less-qualified. A more appropriate title for your article would thus have been "Is there still hope for German universities?" — by dropping the unrealistic and unproductive "plans for change" before it is too late. Isolation from the international scientific community and lowering of standards in teaching and research would otherwise be inevitable.

E.F. STUMPF

*Institute of Mineralogy
and Petrology,
Mining University,
A-8700 Leoben, Austria*

Prizes for Germany

SIR—When Klaus von Klitzing said (see *Nature* 317, 667; 1985) that he hoped it would not be another 22 years before a West German won a Nobel prize (for physics), he was *not* referring to the prize awarded that long ago to the Munich physicist Mössbauer, but to the Heidelberg physicist J. Hans Jensen (1907–1973), who shared the Nobel prize for physics in 1963 with Maria Goeppert-Mayer, La Jolla (one half of the prize, the other half was awarded to Eugen Wigner, Princeton). Rudolf Mössbauer, München, shared the Nobel prize for physics 1961 with Robert Hofstadter, Stanford.

ROSWITHA SCHMID

*Naturwissenschaftliche Rundschau,
Postfach 40,
D-7000 Stuttgart 1, FRG*

UK physics

SIR—I read "Sociological problems of high-energy physics" (*Nature* 318, 243; 1985) with interest but with some dismay, as I did not entirely recognize the field in which I work from the text. The article indulges happily in unsupported generalizations, and hypotheses based on speculations by the gurus of long ago. Perhaps from their American viewpoint they do not see the vitality and innovation of the field over here. Nobody looking at the L3 and DELPHI experiments at LEP could consider them as conservative. Even a project such as OPAL, founded on the premise that it must work reliably from turn-on, is benefiting every day from the individual initiative and innovation of its younger members.

The "10–25 per cent heterodox research" advocated in the article is thriving in the United Kingdom (see the 1984 Rutherford Annual Report for a full list) but will have to diminish if the cuts recommended in the Kendrew report are inflicted.

There is a sociological problem in high-energy physics, certainly in the United Kingdom, forced on us by the persistent negativism of our paymasters. First-rate researchers leave the field, for lack of career advancement, and young people are discouraged from entering as the number of posts is whittled down. (The breathing space afforded by the Science and Engineering Council's New Blood schemes is now over.) This is occurring at a time when the country desperately needs scientists who have been trained in making their mark in a European environment. British industry is not going to be helped by seeking artificially sheltered conditions in which to train our best researchers.

It is fashionable to decry high-energy physics now. Let us not be misled. In spite of official discouragement, the field remains a vital and exciting one in which to work.

D.H. SAXON

*16 Membury Way,
Grove, Wantage,
Oxfordshire OX12 0RP, UK*

AFRC cuts

SIR—Over the past few years you have given detailed consideration to the financial problems of the Agricultural and Food Research Council (AFRC), culminating in your report of the severity of the latest cuts at this institute (*Nature* 317, 663; 1985).

It is not, however, generally appreciated that the extensive cuts at the Institute for Research on Animal Diseases include the savaging of the molecular biology department, which was set up some four years ago to introduce genetic manipulation and allied techniques to the institute's research programme. In view of the stated commitment of AFRC to a programme of modern biology, one would have expected maintenance or even expansion of this programme.

On the contrary, the number of permanent scientific staff in the department has declined at a faster rate than that of the institute staff as a whole. At a time when molecular biology has expanded elsewhere, it is strange that the opposite should be true of AFRC's restructuring plans at this institute.

G. D. HUNTER

*(Chairman, AFRC Higher Grades Panel,
Institute of Professional Civil Servants)
ARC Institute for Research
on Animal Diseases, Compton,
near Newbury, Berkshire RG21 0NN, UK*

Growth by cannibalization

A new calculation of the process of condensation suggests that more attention should be paid to the correlations between neighbouring droplets.

HERE is as neat a piece of classical physics as one could hope to find. The problem is the apparently simple one of predicting the behaviour of an assembly of droplets such as might be found in a condensing cloud in the atmosphere. Large drops will grow at the expense of smaller drops. The classical problem, of obvious importance in meteorology, is to know how the size distribution changes with time. It is now thirty years since the essentials of the problem were stated and solved by two Soviet physicists, I.M. Lifshitz and V.V. Slyozov. The nub of their result is that, where large droplets grow at the expense of small, the average radius increases with time (t) in proportion to the cube root, $t^{1/3}$.

This cube-root dependence on time of the average radius of the droplets in a cloud seems to be borne out by observation, not merely of the sizes of droplets in an atmospheric water cloud but in other comparable situations, the radii of particles crystallizing from a molten metal, for example. But the other features of the size distribution predicted by Lifshitz and Slyozov are not nearly as well established. And now Dr M. Marder from the University of California at Santa Barbara has been able to correct the older arguments in a particularly interesting way (*Phys. Rev. Lett.* **55**, 2953; 1985).

Because the vapour pressure of a small droplet varies inversely with its radius, small droplets are unstable, and will evaporate rather than grow unless the vapour pressure in the immediate environment is greater than that corresponding to equilibrium. This is why cloud formation requires some mechanism of nucleation, and why the environment of a cloud usually manifests a degree of supersaturation with respect to the equilibrium vapour pressure of bulk water at the temperature of cloud formation. But if those conditions are met, it also follows that large droplets will grow at the expense of smaller droplets.

The practical need is to derive quantitative estimates of the preferential growth of larger particles. The process is dominated by diffusion; a shrinking small drop will not be able to continue shrinking if evaporated water vapour accumulates near the drop, increasing the local vapour pressure above equilibrium value. Similarly, a large drop will not be able to continue growing if it should so denude the local environment of water that the vapour pressure is less than the equilibrium value.

So the rate of growth of a particular droplet may be determined by its immediate past history (which determines whether the immediate neighbourhood is enriched by or depleted of water vapour), as well as by the behaviour of nearby droplets which, by shrinking or growing, may add or subtract water vapour from the environment.

The problem is therefore one in kinetic theory, but with the proviso that equilibrium does not apply. The first successful treatment seems to have been given by the two Soviet physicists. Their argument may be summarized as follows. The simplest case is that in which the medium in which the droplets grow is so massively supersaturated that neighbouring drops have no influence on each other and even the effects of surface tension can be neglected. Then, the rate of growth of a particular drop will be determined by the rate at which water vapour can diffuse to reach its surface. In terms of the radius, R , the rate of change will be the simple function $(D/R) \cdot Q^\circ$, where D is the molecular diffusion constant, Q° is the mixing ratio of the water vapour expressed as the ratio of the volume of the water if it were totally condensed to that of the total space occupied and R enters the equation reciprocally because of the geometry of diffusion onto a small supposedly spherical droplet through the surrounding medium.

A missing ingredient from this extreme approximation has the effect of subtracting from Q° an amount d/R , where d is some quantity proportional to surface tension and the reciprocal of R , appearing as a factor. This accords physically with the circumstance that the effect of surface tension is to reduce the rate of growth of drops, the smaller the droplet the more markedly.

The other obviously missing ingredient is some allowance for the extent to which one drop influences its neighbours. First, it is necessary to adjust Q° (by subtraction), the volume fraction of the water already condensed, and then to allow for the way in which the growth of a particular droplet is affected by all the behaviour of all the other droplets in the system, by a separate adjustment representing the influence of each of them. Shrinking droplets will add to the availability of material elsewhere, which explains why the rate of change enters linearly in each adjusting term, adding to the availability of material (and thus Q° when it is negative). The

radius of every other drop, or rather its square, also enters, but the distant effect of each droplet is attenuated in the same way as diffusive effects are invariably attenuated by gaussian factors in respect of both distance and time; in short, the growth of a particular droplet will be more affected by the recent behaviour of other nearby droplets than by the remote (in time) behaviour of distant droplets. The result is a horrid set of nonlinear interlocking equations that nobody would think of solving directly, not merely because solution would be impossible but because the problem is in any case statistical.

Lifshitz and Slyozov dealt with the problem by averaging, on the specific assumption that drops of different random sizes are randomly distributed through the available space. Apart from their conclusion about the average size r , they also concluded that the size distribution of growing droplets, strictly a function both of radius and of time, should ultimately be given by a calculable function only of the ratio R/r , which of course involves the time as a variable because the average radius r grows with time. Marder's point is that while the cube-root dependence of r has been well confirmed by observation, the size distribution of particles that grow partly by the remote cannibalization of each other usually turns out to be much broader than that calculated.

What can have gone wrong? Marder's answer is simple; Lifshitz and Slyozov had no business to determine the average behaviour of a droplet by assuming that droplets of different sizes may be randomly distributed. On the contrary, Marder says, precisely because a shrinking drop will add preferentially to the growth of its immediate neighbours rather than more distant drops, there will be a tendency for large and small drops to be paired together statistically. By good fortune, the more exact correlation is calculable. The proof of the conclusion is that the calculated size distribution agrees well with data previously gathered for the particle-size distribution in slowly solidifying alloys of Fe-Si-Ti and Ni-Si, both of which are broader than predicted by Lifshitz and Slyozov. More detailed calculations of the evolution of a system of mutually cannibalizing particles can pass through a stage in which there is a distinctly bimodal distribution of size, with one group of smallish particles and another which is much larger.

John Maddox

Sex chromosomes

Mammalian X and Y crossover

from Paul S. Burgoyne

HYPOTHESIS and fact, although semantically distinct, are less easy to separate in the real world. Scientific concepts often seem to weave to and fro across this semantic divide before eventually being universally accepted or rejected. A paper by F. Rouyer *et al.*¹ on page 291 of this issue, together with three others recently published in *Nature*²⁻⁴, present incontrovertible evidence for regular exchange (crossing over) between X and Y chromosomes in both mouse and man. This finally validates the concept of X-Y crossing over which dates back to Darlington in 1934, and must surely persuade even the most persistent doubters^{5,6}.

In 1982, two papers^{7,8} appeared that reiterated the case for genetic homology and crossing over within the X-Y pairing segment in man and other mammals. In my paper⁷, I chose to emphasize a conclusion, derived from the unusual inheritance of the sex-reversed factor (*Sxr*) in the mouse, that a single 'obligatory' crossover occurred in the X-Y pairing segment. To describe genes that are exchanged between the X and Y by this crossover and therefore misleadingly seem to be linked to an autosome rather than the X or Y chromosome, I suggested the term 'pseudoautosomal'. Polani⁸ had the foresight to see how molecular genetic approaches could be used to provide the definitive evidence and a search for DNA sequences shared by human X and Y chromosomes ensued. Such sequences were found but, obtusely, they were not located in the X-Y pairing segment. A substantial amount of the X-Y sequence homology proved to be the result of an evolutionarily recent transposition of X-chromosomal material to the Y chromosome⁹.

A non-molecular approach provided the first genetic evidence that crossing over is a regular feature of normal X and Y chromosomes: Keitges *et al.*⁴ demonstrated that a mouse steroid sulphatase (*STS*) variant, which had seemed to be inherited autosomally, was in fact transmitted to XO offspring via the X chromosome. This observation, together with the existing pedigree data, proved that the mouse *STS* gene was present and expressed on the X and Y chromosomes and that there must be an obligatory crossover proximal to the *STS* locus to account for the absence of either X or Y linkage of *STS*. This pseudoautosomal mode of inheritance of the mouse *STS* gene contrasts with that in the human, where the *STS* gene is X-linked, although close to the pseudoautosomal region.

The molecular approach has now also borne fruit with the description of pseudoautosomally inherited human DNA sequences^{2,3}, culminating in the elegant linkage analysis of three pseudoautosomal loci reported by Rouyer *et al.*¹. The features of the obligatory X-Y crossover model⁷, incorporating the most recent findings, are illustrated for the human X-Y pair in the figure. The existence of an obligatory X-Y crossover, at least in those spermatocytes generating functional sperm, is clearly confirmed by X-Y recombination data^{1,4}. This brings the X-Y pair in line with the autosomes, where there is a minimum of one chiasma, the morphological manifestation of crossing over, for each meiotic chromosome pair (bivalent), irrespective of the size of the chromosome.

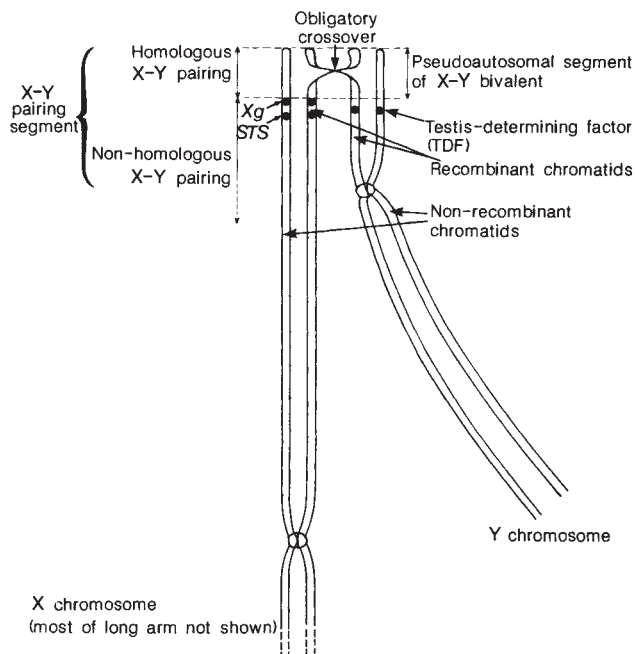
Although in most mammals no obvious X-Y chiasma can be seen by light microscopy, chiasmate X-Y associations have been demonstrated with the electron microscope¹⁰. They have not been found in all X-Y bivalents but this could be due to technical difficulties. In some instances the absence of an X-Y chiasma may result from inadequate X-Y pairing. In XY *Sxr* mice there is a high incidence of X-Y separation at pachytene, presumably because the presence of *Sxr* distal to the pairing segment sometimes interferes with the pairing process. Nevertheless, the recombination data for *Sxr* confirm that all functional sperm are derived from spermatocytes that had an X-Y chiasma, adding weight to the view that there is a meiotic 'quality-control' mechanism that prevents primary spermatocytes with unpaired chromosomes from forming viable sperm¹¹.

I find that it helps to think of the pseudoautosomal segment of the X-Y bivalent as a very small extra pair of autosomes. Rouyer *et al.*¹

estimate that the human pseudoautosomal segment contains 2,000–5,000 kilobases, which would make it about one-tenth the size of the smallest human autosomal bivalent. An obligatory crossover in such a short length inevitably results in a high recombination frequency per unit length¹. In the X bivalent of females, the obligatory crossover is not constrained within the pseudoautosomal region, so recombination in this region is reduced almost tenfold¹.

What determines the limits of the pseudoautosomal segment? As originally defined⁷, it is that part of the X-Y bivalent where there is exchange by crossing over, a definition that includes genes showing partial sex linkage — those with less than 50 but greater than 0 per cent X-Y recombination.

In practice it may be necessary to set an arbitrary minimum of say 1 per cent recombinants as defining a pseudoautosomal locus, as it is important to exclude those X-linked (for example, *Xg*) or Y-linked (for example, *TDF*) genes that are occasionally exchanged between the X and Y chromosomes¹² by unknown mechanisms



Crossing over between the human X and Y chromosome. The pseudoautosomal segment is that part of the X-Y bivalent where there can be X-Y exchange by crossing over. X-Y homology in this segment is maintained by, and may be necessary for, this crossing over. There is always one 'obligatory' X-Y crossover but, because its position varies, only the more distal pseudoautosomal loci show 50% X-Y recombination. More proximal loci show a hierarchy of partial sex linkage (<50% but >0% X-Y recombination), reflecting their order within the pseudoautosomal segment¹. X-dosage compensation is unnecessary for pseudoautosomal genes and they are expressed in double dose in both males and females. X-linked genes bordering the pseudoautosomal segment (*Xg*, *STS*) are also at least partially expressed on the 'inactive' X of females. The length of the X-Y pairing segment varies with meiotic stage and can extend well beyond the pseudoautosomal segment into the Y long arm. Much (perhaps all) of the synaptonemal complex formed outside the pseudoautosomal segment represents non-homologous pairing. (Modified from ref. 7.)

that are not formally distinguishable from crossing over.

These problems of definition apart, what is it that constrains the obligatory crossover to occur in the pseudoautosomal segment of the X-Y bivalent? The constraint is certainly not the length of the pairing segment — in man, X-Y pairing, as defined by the formation of a synaptonemal complex, includes most of the Y short arm and can extend into the long arm, much longer than the pseudoautosomal segment. More probably it is the extent of X-Y homology in the pairing segment.

In other situations where stretches of synaptonemal complex are formed between non-homologous chromosome segments, crossing over is excluded from these non-homologous regions. Because there are degrees of homology, the pseudoautosomal segment may end in a zone where the likelihood of a crossover rapidly decreases in conjunction with a decrease in homology, rather than having a discrete boundary.

It is generally agreed that the heteromorphic mammalian X-Y pair must have evolved from a pair of homologous chromosomes. The pseudoautosomal segment might therefore represent a region of the original homologous pair that has been retained to ensure chiasma formation and orderly X-Y separation at meiotic metaphase I. Although gene-dosage considerations do not preclude the pseudoautosomal segment from exchanging material with the autosomes¹, it may nonetheless be a conserved linkage group with significant homology between mammalian species. If so, with relaxed hybridization conditions it should be possible to use human pseudoautosomal DNA probes to pull out related pseudoautosomal sequences from the mouse genome.

Combining the flexibility of mouse breeding with the exceptionally high recombination frequency in the pseudoautosomal segment should allow a rapid and detailed genetic analysis of this region. Now the molecular geneticists have stuck in their thumbs, I am sure that they will pull out some plums. 1986 promises to be a good year. □

1. Rouyer, F. *et al.* *Nature* **319**, 291 (1986).
2. Cooke, H.J. *et al.* *Nature* **317**, 687 (1985).
3. Simmler, M.-C. *et al.* *Nature* **317**, 692 (1985).
4. Keitges, E. *et al.* *Nature* **315**, 226 (1985).
5. Ashley, T. *Hum. Genet.* **67**, 372 (1984).
6. Ashley, T. *Genetica* **66**, 161 (1985).
7. Burgoyne, P.S. *Hum. Genet.* **61**, 856 (1982).
8. Polani, P.E. *Hum. Genet.* **60**, 207 (1982).
9. Page, D.C. *et al.* *Nature* **311**, 119 (1984).
10. Holm, P.B. & Rasmussen, S.W. *Carlsberg Res. Commun.* **48**, 415 (1983).
11. Burgoyne, P.S. & Baker, T.G. in *Controlling Events in Meiosis* (eds Evans, C.W. & Dickenson, H.G.) **349** (Company of Biologists, Cambridge, 1984).
12. De la Chapelle, A. *et al.* *Nature* **307**, 172 (1983).

Paul S. Burgoyne is in the Medical Research Council Mammalian Development Unit, University College London, 4 Stephenson Way, London NW1 2HE, UK.

Nuclear winter

Towards a scientific exercise

from K.A. Emanuel

THE detonation of the first atomic bomb on 16 July 1945 changed the world in many ways, not least of which is the view some scientists have taken towards their profession. Many have come to feel that responsibility for the influence of their achievements rests partially with them and should not reside solely with politicians. While imparting certain benefits to political decision making, this attitude has at times tainted the objectivity that is crucial to scientific endeavour. Nowhere is this more apparent than in the recent literature on 'nuclear winter' research, which has become notorious for its lack of scientific integrity. Among the most serious criticisms levelled at this work has been the failure to quantify the large uncertainties associated with estimates of the war-initiated fires and their combustion products, the highly approximate nature of the global circulation models used in the calculations, and the appearance of the results in popular literature before being exposed to the rigours of peer review.

Although controversy continues to surround this work, serious research is beginning to qualify the earlier bold assertions. A good example appears on page 301 of this issue in the paper by B.W. Golding, O. Goldsmith, N.A. Machin and A. Slingo, which attempts to examine the physical factors affecting the disposition of a smoke plume shortly after its inception. These authors point out that previous estimates of the nuclear winter effect have relied either on one-dimensional models, which must assume a uniform distribution of smoke, or on three-

dimensional models with grid intervals too large to resolve mesoscale (medium-scale) circulations very near recently initiated plumes. Yet it is likely that such circulations would result from the horizontal temperature gradients produced by radiative heating of the smoke.

Using a mesoscale numerical model, Golding *et al.* show that mesoscale ascent rates of 20 cm s⁻¹ are possible 12 h after plume initiation; such rates will rapidly lead to condensation of water vapour even in circumstances where the air is initially quite dry. Subsequent scavenging of the smoke by precipitation could substantially alter the characteristics of the plume before it diffuses into the continental-scale pall which is the starting point of most nuclear winter calculations.

The calculations presented by Golding *et al.* are themselves dependent on assumptions concerning the concentration, geometry and radiative characteristics of the smoke plume and the large-scale meteorological conditions in which it is released, and the model does not explicitly deal with the interaction of condensed water and smoke particles. The results do strongly suggest, however, that more attention must be paid to the mesoscale characteristics of the plume in the first day or so of its lifetime. The paper is a welcome step in transforming nuclear winter research from a means of political advocacy to a scientific exercise. □

K.A. Emanuel is at the Center for Meteorology and Physical Oceanology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Archaeology

Reconstructing ancient diets

from Richard Burleigh

A PAPER published recently (*Nature* **317**, 806; 1985) and one on page 321 of this issue make significant new contributions to the investigation of ancient human diets by the analysis of stable isotopes in skeletal remains. In the first, M.J. DeNiro is concerned with the detection of changes in the composition of sub-fossil bone (diagenesis) that might lead to false results. The other, by S.H. Ambrose and M.J. DeNiro, deals with reconstruction of patterns of diet in humans who are known from historic records or presumed from archaeological evidence to have followed particular modes of subsistence. Both papers help to provide a firmer basis for

possible investigation of human diet in the much more remote past.

Measurement of the ratios of stable carbon isotopes (¹³C/¹²C) in collagen from skeletal remains has been used successfully in recent years to establish prehistoric human diets and hence, somewhat more inferentially, patterns of subsistence in the past, for example, in parts of North America, Africa and north-west Europe. The basis of this method is that the carbon-isotope composition of collagen in bone reflects the isotopic composition of the plants that form the base of the food web, isotope ratios being expressed as δ¹³C values, or the difference in parts per

thousand from a standard, as measured by mass spectrometry. Thus, the relative importance in the diet of C_3 and C_4 plants, which have different photosynthetic pathways characterized by distinct $\delta^{13}C$ values, can be determined. Marine sources of food can confuse this picture but more recently it has been found that, by simultaneous measurement of stable nitrogen ratios ($^{15}N/^{14}N$), the respective contribution of marine and terrestrial sources of food can be distinguished from the differences in $\delta^{15}N$. Because there is a small incremental enrichment of ^{13}C and ^{15}N as these isotopes are moved up the food chain, the $\delta^{15}N$ values in particular also give an indication of the relative importance in the diet of animal products compared with more direct exploitation of plant foods.

Ambrose and DeNiro have investigated the $\delta^{13}C$ and $\delta^{15}N$ values of collagen recovered from the bones of about 100 individuals representative of particular groups of hunter-gatherers, pastoralists and agriculturalists both from the historic period and from archaeological sites in Kenya, Tanzania and South Africa. The bones of wild mammalian species representing browsers feeding mainly on C_3 trees, of grazers feeding mainly on C_4 grasses and of carnivores, also analysed, showed isotopic differences clearly reflecting these animals' diets (with an apparent preference by carnivores for preying on grazers). These results provide a convincing demonstration that in the same biome in Africa, human diets based variously on C_3 and C_4 plants, meat of grazing and browsing animals, and other animal products such as milk and blood, should be distinguishable by a combination of the $\delta^{13}C$ and $\delta^{15}N$ values.

The diets of the historic populations investigated were known from ethnographic records, and Ambrose and DeNiro found very close agreement of the isotope values with this evidence. This in turn gives confidence that their reconstruction of the diets of the prehistoric groups from isotope analyses is essentially correct, as direct archaeological evidence of food resources in the form of animal and plant remains is inevitably incomplete and may sometimes be misleading.

A logical step from this is, the authors suggest, the intriguing possibility of reconstructing the diets of much earlier human ancestors. The living primates most closely related to *Homo sapiens* are predominantly C_3 plant feeders. Australopithecines probably were as well, but there is evidence to suggest that the true hominids, *H. habilis* and *H. erectus*, were more dependent on meat. Independent validation of this could have great importance for the study of human evolution, but depends entirely on the isotope record being preserved in bones of Pliocene-Pleistocene age, that is, at least 4 million

years old. This may be unattainable in practice, though the photosynthetic pathways with their isotopic differences were certainly established well before then.

A more immediate question is raised by DeNiro, who examines the assumption, implicit in most previous studies, that the isotopic composition of collagen in the bones has not subsequently been adversely altered. Inevitably sub-fossil bone will have undergone physical and chemical changes, which may have resulted in the addition or removal of organic components; much therefore depends on the extent to which the organic fraction used for analysis represents the original collagen. To test this, he analysed bones of marine carnivores and terrestrial herbivores from archaeological sites varying in age from hundreds to thousands of years, and from various depositional environments ranging from arid desert conditions to peat bogs. These species were chosen because the feeding patterns and isotope values of their modern counterparts have previously been well established.

DeNiro finds that, independent of age, those sub-fossil bones with C/N ratios in the range 2.9 to 3.6, characteristic of fresh modern bone, have $\delta^{13}C$ and $\delta^{15}N$ values consistent with expectation. In contrast,

burnt bones, where the C/N ratios are outside the normal range, have isotope values that are shifted erratically. It is a reasonable assumption that at least some of the other, often individually indeterminate, diagenetic processes to which sub-fossil bone is subjected might produce similar isotopic shifts. A sensitive criterion, representing the sum of these processes, that allows altered material to be easily recognized and rejected in advance is therefore of the greatest value, although it cannot be applied retrospectively to previous studies without published C/N ratios.

The determination of stable carbon and nitrogen isotope ratios also has purely biological applications, for example, to the investigation of the differential feeding patterns of closely related animal species living in restricted environments, such as those found on remote oceanic islands. These two new papers not only add rigour to a scientific method that is beginning to give results of real importance to archaeologists and anthropologists but are also highly relevant to the refinement of other applications of the method. □

Richard Burleigh was at the Research Laboratory, The British Museum, London WC1B 3DG, and is now at the Museum of Mankind, 6 Burlington Gardens, London WC1X 2EX, UK.

Ecology

Tundra river transformation

from S.I. Heaney

THE past quarter century has seen a great interest and concern over anthropogenic phosphorus loading of lakes and reservoirs. Such enrichments of standing waters often lead to the formation of troublesome surface blooms of algae and deoxygenation of the colder lower layers (Edmondson, W.T. & Lehman, J.T. *Limnol. Oceanogr.* **26**, 1; 1981). Much less attention has been paid to the role of phosphorus in running waters. So the report by B.J. Peterson *et al.* (*Science* **229**, 1383; 1985) on the effects of experimental addition of phosphorus to a nutrient-impooverished arctic river is both important and surprisingly overdue.

Peterson's group added phosphorus to a portion of the Kuparuk River in the Alaskan tundra, raising the concentration of PO_4-P from ambient levels of $< 1-4 \mu g l^{-1}$ to $10 \mu g l^{-1}$, and then monitored the river upstream and downstream of the site of addition. Within six weeks the biomass of the algae (measured as μg chlorophyll *a* per cm^2) attached to the stones on the river bottom had increased tenfold within a kilometre of the site of phosphorus addition. This increased biomass was sustained for several kilometres downstream but there was a negligible change of biomass upstream. The result is probably

to have been expected considering the concentrations of other dissolved plant nutrients (carbon, nitrogen) in the stream water. However, the strength and originality of the work lies just as much in the efforts made to separate and quantify the microbial metabolism into heterotrophic (bacterial) and autotrophic (algal) processes and to attempt to see how these induced microbial changes influenced the higher trophic levels (dominant aquatic insects) of the ecosystem.

No less than eleven authors from four institutions were involved in the study, which involved several disciplines and techniques and made possible a holistic approach to this relatively simple ecosystem. Although the results are necessarily restricted they appear convincing enough to support the conclusions: first, photosynthesis and biomass production in the river are limited by phosphorus concentration. Second, on the addition of phosphorus, photosynthetically fixed energy replaced organic matter entering the stream as the major energy source. Third, as a result of increased photosynthesis, algal activity increased on rocks of the river bottom. And fourth, as illustrated by the increase in the size and developmental stage of some aquatic

insects, there was food limitation at higher trophic levels in this nutrient-poor environment.

This innovative experiment provokes further questions. Phosphorus was added only during July and August when water temperatures were between 7 and 14°C. As some freshwater algae, including diatoms, can grow at temperatures down to freezing with enough light and nutrients, what would be the effect on the community if phosphorus were added throughout the ice-free period? With continuously elevated phosphorus concentrations, what would be the long-term consequences on the abundance and periodicity of algae other than diatoms? Is the timing of the spring algal growth period significant in relation to the emergence and growth of dominant insect grazers? What is the

relationship between suspended phytoplankton and attached algae?

Answers to some of these questions can be sought from studies of phosphorus-enriched rivers. Nevertheless, the experimental work of Peterson's group, earlier pioneer work in experimentally increasing heterotrophic microbial production in trout streams in Oregon (Warren C.E., Wales, J.H., Davis, G.E. & Doudoroff, P. *J. wildl. Mgmt.* **28**, 617; 1964) and that of the Freshwater Biological Association using its experimental circulating channels (Marker, A.F.H. & Casey, H. *Phil. Trans. R. Soc. B* **298**, 265; 1982) will be vital for a fuller understanding of the processes controlling production in rivers. □

S.I. Heaney is at the Freshwater Biological Association Windermere Laboratory, Ambleside, Cumbria LA22 0LP, UK.

Developmental biology

Unscrambling egg structure

from Hugh Woodland and Elizabeth Jones

In these columns, we recently reviewed the evidence for the suggestion that the complexity of the amphibian embryo is generated by progressively more complex interactions between a few cell types in the early embryo¹. These cell types, possibly just the ectoderm and endoderm, probably arise by the partitioning of different cytoplasmic constituents from the egg. What could these differences be?

In the amphibian, one of the few ways to identify them is to pick a molecule and look for its uneven distribution in the egg. There is some evidence from other organisms, particularly insects, that RNA which specifies future cell types is distributed in this way. Several groups have now looked to see if particular messenger RNAs show differences along the animal-vegetal axis of *Xenopus* eggs. *In vitro* translation has shown that at least 17 mRNA species are asymmetrically distributed, which is 4 per cent of the total translatable RNA².

The logical extension of this approach is to clone such mRNAs; this has now been achieved for four mRNAs which are concentrated in either the animal hemisphere (An1-3) or in the vegetal hemisphere (Veg1)³. All these RNAs are accumulated steadily during oogenesis and two of them, An1 and Veg1, are synthesized only during oogenesis and disappear during gastrulation. The behaviour of these transcripts, especially An1 and Veg1, suggests that they have an interesting role in development.

The ease with which antibodies, as well as sense and anti-sense RNA, can be injected into specific cells in the cleaving embryo suggests obvious approaches to finding their function. Sequencing the clones also suggests a function for the

mRNAs; it would be particularly interesting to find hints of activities that could set up positional information (such as gradients of small molecules) to provide the spatial coordinates by which developing cells sense their position and respond accordingly.

Drosophila

If one employs genetics to approach this same subject, one can at least start with a knowledge of what the genes do in the developing animal, because they will have been identified by virtue of some interesting, probably dramatic, effect on the developing embryo. This is why current intense activity in *Drosophila* developmental genetics is so exciting for embryologists of all kinds. Most such work recently published in these pages has been concerned with homeotic genes, like those of the antennapedia complex, or the segmentation genes, for example, *fushi tarazu* and *engrailed*. These genes are expressed after fertiliza-

tion and interpret spatial cues established at earlier stages. The latter are produced by other genes transcribed during oogenesis, that is, they are 'maternal effect' genes. These genes do not affect the maternal carrier, but their malfunctions are revealed as abnormalities in all eggs produced by females carrying any copies of the mutant gene in the case of dominant genes, or lacking wild-type genes in recessives. Such genes have global effects on the major body axes of the embryo, to be expected if they set up some kind of network of positional information, such as gradients.

Judging by the major classes of maternal mutants discovered, there seems to be two kinds of positional information in the early insect embryo, that for the antero-posterior axis and that for the dorso-ventral axis. These two sets of information would be sufficient to make every point in the left or right halves of the embryo unique. Our knowledge of the dorsal group of genes has mainly derived from the work in Christiane Nüsslein-Volhard's laboratory⁴⁻⁶. At least 10 separate loci are involved in establishing the dorso-ventral axis, all manifesting themselves as recessive mutations in which the developing embryo lacks cells of the ventral type. Because a recessive mutation usually results in the loss of function of a gene (for example, its deletion, or encoding of a non-functional protein) this indicates that the products of these genes are necessary to form ventral structures (their confusing name of 'dorsal-type' is because genes are typically named after the observed effect of their malfunction, and not that of their normal function). A particularly important feature of most of these genes is that they can be rescued by injecting cytoplasm into the eggs of mutant females, thus restoring the normal phenotype at least to some extent. In all cases where this can be achieved except one (*dorsal*), pure RNA is also effective (see table).

Two recent reports^{7,8} indicate that one of these maternal effect genes, *Toll*, is particularly interesting. *Toll* is unique among such genes because it not only has recessive, lack-of-function dorsalizing alleles, but also has dominant gain-of-function ventralizing alleles. Perhaps the most immediately striking observation about all the dorsal-type mutants is that among the mutants rescuable by injection with RNA, it is only in *Toll*⁻ embryos (those laid by females homozygous for *Toll*⁻ recessive mutations) that it matters where the cytoplasm or RNA is injected⁸. The point of injection marks the future ventral region of the rescued embryo. In the other dorsal-type mutants, the ventral region is always in the usual place, recognizable by the asymmetrical shape of the egg. This indicates that the dorso-ventral polarity still exists in some form in these embryos,

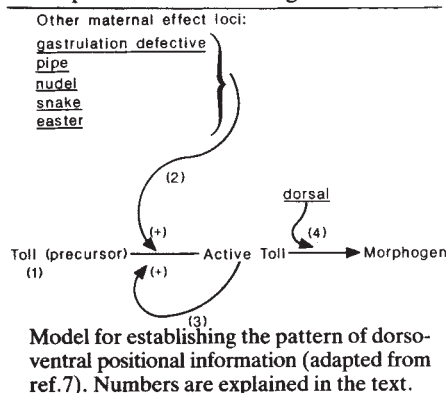
Maternal effect loci affecting the dorso-ventral axis (from refs. 6 and 7).

			+ type Rescue	
			Cyto- RNA	
			plasm	
<i>Gastrulation defective</i>	Recessive	Dorsalized	-	ND
<i>Pipe</i>	Recessive	Dorsalized	-	ND
<i>Dorsal</i>	Recessive	Dorsalized	+	-
<i>Nudel</i>	Recessive	Dorsalized	-	ND
<i>Tube</i>	Recessive	Dorsalized	+	+
<i>Snake</i>	Recessive	Dorsalized	+	+
<i>Easter</i>	Recessive	Dorsalized	+	+
<i>Toll</i>	Recessive	Dorsal	+	+
	Dominant	Lateralized or ventralized	ND	ND
<i>Spätzle</i>	Recessive	Dorsalized	+	+
<i>Pelle</i>	Recessive	Dorsalized	+	+

but needs the expression of the non-*Toll* dorsal-group genes to be interpreted. *Toll*⁻ embryos must have no inherent dorso-ventral axis, and at least a partial axis is organized at the position where normal cytoplasm is injected.

Anderson, Jürgens and Nüsslein-Volhard have conducted an intensive genetic analysis of *Toll*, which has led them to postulate that the *Toll*⁺ product is immediately involved in setting up the hypothetical morphogen gradient that establishes the dorso-ventral axis⁷. It cannot be the morphogen itself because rescue of the eggs of *Toll*⁻ females can be achieved with cytoplasm taken from anywhere in the wild-type egg (see figure).

Although the *Toll*⁺ product is evenly distributed throughout the embryo, it is presumably only active in the ventral region. Hence it may be present as a precursor which is activated only in this region (1 in the figure). The 'precursor' could be mRNA, in which case 'activation' is its controlled translation. Because *Toll*⁺ RNA produces a ventral region where it is



injected into an egg lacking the *Toll*⁺ product, they propose that the precursor self-activates in the absence of other *Toll*⁺ products, although this explanation is not very satisfactory without further evidence.

There are four dominant alleles of *Toll*, divided into two groups. In eggs from heterozygous females carrying all these mutations, only ventral structures are present. When one of the pair of alleles *Tl*^{5b} and *Tl*^{84c} is present together with any homozygous dorsal-type recessive mutation from those listed in Table 1, the resulting embryos have only dorsal structures. Thus, *Toll* requires each one of these dorsal-type gene products for its activity (2 in the figure). Furthermore, these dominant alleles *Tl*^{5b} and *Tl*^{84c} only produce their totally ventral phenotype when there is also a copy of *Toll*⁺ present, that is, in the absence of normal *Toll*⁺ product they produce the totally dorsalized phenotype characteristic of no *Toll* activity, as seen in the recessive *Toll* mutants. This could mean that *Toll* produces a multi-subunit protein, which needs some normal subunits for these dominant alleles to exert their effect. However, the authors favour the possibility that the active form

of the product stimulates its own synthesis, *Tl*^{5b} and *Tl*^{84c} having lost this function.

The other two dominant alleles of *Toll* (*Tl*⁺ and *Tl*⁹⁰) will at least partially override all of the dorsal-type genes, except *dorsal* itself, to produce intermediate lateralized embryos. In addition, these alleles do not need the presence of a normal *Toll* gene for their expression. It is therefore assumed that *Tl*⁺ and *Tl*⁹⁰ bypass the precursor activation step by making an altered *Toll* product that is active or self-activating, hence uncoupling the dependence of *Toll* on the other dorsal-type genes and eliminating the positive-feedback loop (3 in the figure). The role of the dominant alleles of *Toll* in interactions with other dorsal-type genes has also been investigated by cytoplasmic rescue experiments⁸. The cases of *gastrulation defective* and *pipe* are especially interesting, because eggs affected by mutations in these genes cannot be rescued by the injection of normal cytoplasm. This can perhaps be explained for *gastrulation defective* as it has a temperature-sensitive period (time of activity) from before to a little after fertilization, earlier than the other genes tested. Thus, cytoplasm is injected into these eggs too late to effect rescue. However, cytoplasm from the *Tl*⁹⁰ dominant eggs will rescue them, emphasizing the escape of the *Tl*⁹⁰ product from control by the other genes.

All the *Toll* dominant alleles require the presence of an active *dorsal* gene in the female to produce their ventralizing effect (4 in the figure). This suggests a special position for the *dorsal*⁺ product. It is put later in the chain of events because rescue by *dorsal*⁺ cytoplasm requires the injection to be ventral, that is, into the area where the hypothetical *Toll* product is activated. *Dorsal*⁺ has an early time of action, overlapping but extending longer than *gastrulation defective*. It can only be rescued by cytoplasm and not by RNA, suggesting that it is translated early on. However, it is a puzzle that the *dorsal* RNA is actually present throughout early development¹⁰. Validation of this interesting model will require the cloning of the genes concerned, which is no small task.

It is intriguing that a new example of a maternal effect gene has been isolated by virtue of its containing a homoeo box sequence¹¹. This gene, *caudal*, is unknown genetically but is expressed in oogenesis and its product becomes localized to a band at the posterior tip of the embryo at the cellular blastoderm stage. This gene may be a relative of the bithorax complex controlling the development of the partially suppressed ninth abdominal segment. The most interesting feature of *caudal* is that it is the first maternally expressed *Drosophila* gene with a homoeo box, although one such *Xenopus* gene also has this sequence¹². This suggests a new way to obtain genes that control

100 years ago



THE Lions' House in the grounds of John Hunter's house at Earl's Court. It was "a raised mound of earth. The earth rests upon an arched structure, which, at the time of my visit, was in excellent condition." The sketch was one of a series of homes and birthplaces of illustrious men drawn by Bertram Richardson.

from *Nature* 33 275, 21 January 1886.

patterns of vertebrate development, and re-emphasizes the considerable general impact of *Drosophila* genetics.

Finally, there are two parallels which may be drawn between the establishment of the dorso-ventral axis in insects and vertebrates. First, the key area in the dorso-ventral positional field of insects is the ventral part of the embryo. This is where the mesoderm develops, and the region is supposed to organize the dorso-ventrally graded ectoderm dorsal to it by some kind of long-range interaction. In vertebrates, the central nervous system (the most dorsal ectoderm) is also generated by a signal from the mesoderm, the phenomenon called neural induction¹. Second, in insects the ventral *Toll*⁺ component may be generated by activation of an inactive precursor (or translation of an RNA). It is proposed that the amphibian dorso-ventral axis is also generated by the unmasking of a ubiquitous component, although here it marks the dorsal limit of the dorso-ventral axis and the response is polarized by a cytoplasmic movement, usually in response to sperm entry¹³. The parallels are certainly intriguing and it will be fascinating to see how real they are. □

1. Woodland, H.R. & Jones, E.A. *Nature News and Views* 318, 102 (1985).
2. King, M.-L. & Barklis, E. *Dev. Biol.* 112, 203 (1985).
3. Rabagliati, M.R., Weeks, D.L., Harvey, R.P. & Melton, D.A. *Cell* 42, 769 (1985).
4. Nüsslein-Volhard, C. in *Determinants of Spatial Organization* (eds Subtelney, S. & Konigsberg, I.R.) 185 (Academic, New York, 1979).
5. Anderson, K.V. & Nüsslein-Volhard, C. *Nature* 311, 223 (1984).
6. Anderson, K.V. & Nüsslein-Volhard, C. in *Pattern Formation* (eds Malacinski, G.M. & Bryant, S.) 269 (Macmillan, New York, 1984).
7. Anderson, K.V., Jürgens, G. & Nüsslein-Volhard, C. *Cell* 42, 779 (1985).
8. Anderson, K.V., Bokla, L. & Nüsslein-Volhard, C. *Cell* 42, 791 (1985).
9. Santamaria, P. & Nüsslein-Volhard, C. *EMBO J.* 2, 1695 (1983).
10. Steward, R., McNally, F.J. & Schedl, P. *Nature* 311, 262 (1984).
11. Meodzik, M., Anders, F. & Gehring, W.J. *EMBO J.* 4, 2961 (1985).
12. Müller, M.M., Carrasco, A.E. & DeRobertis, E.M. *Cell* 39, 157 (1984).
13. Gerhart, J.C. et al. *Proc. R. Soc. B* 307, 319 (1984).

Hugh Woodland and Elizabeth Jones are in the MRC Developmental Biology Group, Department of Biological Sciences, Warwick University, Coventry CV4 7AL, UK.

Solar physics

What causes the solar cycle?

from Douglas Gough

THE dynamics of the solar cycle are not understood. The main obstacle to progress is the paucity of data in a form that is immediately pertinent to the alternative theories that have been put forward to account for the cycle. That situation is bound to be improved by the interesting new analysis of solar magnetic fields reported by J. O. Stenflo and M. Vogel on page 285 of this issue¹.

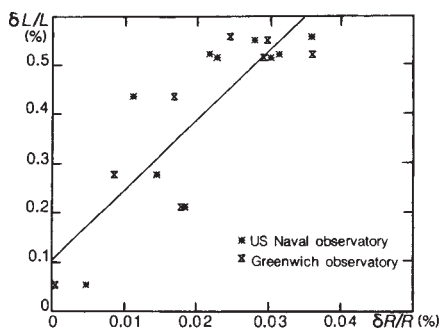
The most popular theories presume that the cycle results either from a coherent magnetically controlled global oscillation of the Sun, or from some dynamo process that is essentially confined to the turbulent convection zone and possibly its immediate environs. The first alternative is perhaps the more obviously appealing, for it yields the observed 22-year period in a quite natural way, requiring simply that there is a large-scale internal magnetic field whose intensity is comparable to that observed in sunspots. Dynamo theories are less transparent, but they can rationalize the observed period by adjusting the quite uncertain statistical properties of the turbulence. They have been more highly developed, and consequently have gathered a larger following. But their failure to explain, in particular, the migration towards the equator of the sunspot belts has cast doubts even in the minds of some of their most ardent supporters. There is also evidence that the mechanism of the cycle extends well beneath the base of the convection zone.

One of the first hints that the solar cycle is not confined to the convection zone came from a dendrochronological analysis by Stockton *et al.*² of drought in North America. It was anticipated there would be a correlation with the solar cycle, with a period of 11 years. If contemporary dynamo models were correct, the entire magnetic field would be the same every half-cycle. Because the structure of the Sun cannot depend on the polarity of the field, any solar variation that might influence the climate should therefore have the 11-year period.

But to everyone's surprise, the drought period was found to be 22 years. Therefore the Sun can sense the difference between its north and its south, which can be so only if there is a nonreversing field, presumably in the radiative interior, that is strongly coupled to the oscillating field outside. It has been claimed that the Princeton oblateness measurements³ and the fine structure in the power spectrum of low-degree acoustic oscillations⁴ also show indirect evidence for a deep-seated magnetic field. Moreover, the existence of

widely spaced stable longitudes of magnetic activity seems to be more readily comprehensible if there is a coherent field that extends into the radiative interior.

The climatic record could be regarded as no more than a clue, for perhaps the solar-terrestrial connection filtered the 22-year period: if climatic variation results from the modulation of the Earth's ionosphere by charged particles from the Sun, it could be that the geomagnetic field determines the invariant direction. However, corroborating evidence that the direction is determined in the Sun was provided by Sonnet⁵, who pointed out a systematic difference between odd and



Relation between relative radius ($\delta R/R$) and solar luminosity ($\delta L/L$) variations in the period 1967–1982 (after ref. 6).

even sunspot cycles. And more recently, Eddy and Fröhlich⁶ have found decade timescale variations in 17-year records of solar constant and diameter measurements that seem to be consistent with a 22- but not a 11-year periodicity.

Another interesting result that emerged from the analysis by Eddy and Fröhlich is the ratio, W , of the infrared relative radius and solar luminosity variations (see figure). It had previously been shown that the value of W is an indicator of the depth, d , in the Sun at which the variations are produced⁷ and it had also been conjectured that W is an increasing function of d . Eddy and Fröhlich obtained a value for W ($\approx 0.078 \pm 0.026$) which points to a disturbance near the base of the Sun's convection zone, a natural location for the alternating external field and the direct internal field to meet, rather than in the superficial layers or the energy-generating core. One is reminded of the proposal by Spiegel and Weiss⁸ that the main role of the turbulence in the convection zone is to concentrate the field at its base until instability causes the field to erupt into the start of a new sunspot cycle.

This argument too must be viewed with caution. Variations in the Sun's diameter are notoriously difficult to measure. On

timescales of centuries widely divergent results have been reported. Much of the discrepancy has been attributed to errors introduced by changes in the observing conditions. The data analysed by Eddy and Fröhlich⁹ seem to have been obtained under relatively uniform conditions, so perhaps their variation is a more reliable indicator of a corresponding solar variation. But, they may not reflect gross compressions and rarefactions of the Sun; superficial changes in the scale of solar granulation and in the structure of the solar atmosphere that have been reported to be correlated with the 11-year sunspot cycle may have 22-year counterparts that influence the size of the solar image.

Even if the diameter measurements have been correctly interpreted, one cannot be sure that a process concentrated near the base of the convection zone is responsible for the variations. First, recent calculations^{9,10} have shown that W is not a function solely of d . Second, the seat of the cycle may not be localized near a particular depth; perhaps the value of 0.078 for W is an average response to a mechanism acting throughout the interior of the Sun. Indeed, R. Davis (personal communication), among others, has noticed a correlation between a decade timescale variation in the neutrino flux and sunspot numbers. The result is barely significant but if it does reflect a causal connection, one would surely be compelled to conclude that the core partakes in the cycle. In this connection it is worth noting recent evidence that the core is rotating more rapidly than the rest of the Sun¹¹; the angular velocity variation is unstable, and one might conjecture that the ensuing motion, which is predominantly on spherical surfaces, oscillates with the solar cycle. If there were also a radial advection of material, the variation of chemical composition in the core would modify the neutrino production rate.

Until now there has been only meagre evidence on which to base conjecture but the analysis by Stenflo and Vogel of the Mount Wilson / Kitt Peak magnetic-field measurement might be the start of a new era in solar-cycle diagnosis. Inspired by studies of five-minute oscillations, they projected the signal at each moment on to (axisymmetrical) spherical harmonics and Fourier analysed the resulting time series. Several remarkable properties are revealed, the most surprising of which is that harmonics of odd and even degree, l , behave quite differently.

All odd harmonics with $l < 14$ exhibit the 22-year periodicity. That is to be expected, for it is well known that the oscillating magnetic field is dominated by a component that is antisymmetrical about the equatorial plane; the field is unlikely to be spherically harmonic, so the periodicity should be present in all odd-degree projections. And as the cycle is

commonly considered to be a nonlinear phenomenon, it should not be surprising that no other incommensurable period is present. But it was not anticipated that the coherent variation would be almost sinusoidal, as indicated by the lack of the temporal harmonics of 22 years.

The power spectrum of the even-degree components with $l \leq 14$ is quite startling for, as Stenflo and Vogel point out, it looks superficially like a branch of the spectrum of the five-minute oscillations. Is it a magnetic analogue of linear p modes, as Stenflo and Vogel suggest? If so, why should the spatial structure of the oscillating field be spherically harmonic, for the basic state cannot be spherically symmetrical? Why does there appear to be just one branch of the dispersion relation? Is it only that one branch which has a positive growth rate in linear theory, or do nonlinearities suppress the other modes? And why do only the $l=10$ (and possibly the $l=14$) even-degree modes show power near 22 years? One is reminded of the apparently anomalous rotational splitting of the (neighbouring) $l=11$ acoustic modes reported by Duvall and Harvey¹².

Perhaps the most exciting prospect offered by Stenflo and Vogel's p -mode analogy is the possibility of a seismological inversion to determine the magnetic field inside the Sun. If that prospect becomes reality, many questions about the nature of the solar cycle might be answered. But first we must understand the nature of the magnetic oscillations, and learn how to compute their frequencies. Since, unlike the five-minute modes, the eigenfunctions are not obviously separable, that is likely to be a formidable task, even if linearized theory is adequate. Simple plane-wave analysis such as that presented by Gilman¹³ may provide insight, if not quantitatively correct results. Indeed, such analysis suggests that there are three obvious long timescales that are relevant to the problem: the rotation period P , a characteristic Alfvén time T_A , and the time $\tau = T_A^2/P$. It is interesting that if $\tau=22$ years, then $T_A = (22 \text{ years} \times 25 \text{ days})^{1/2} \approx 1.2$ years, which is close to the period at which the curve in Stenflo and Vogel's Fig. 2 flattens. We must await further research to know whether that correspondence is causal or accidental.

We also eagerly await the analysis of the nonaxisymmetrical modes that Stenflo and Vogel promise. If their p -mode analogy is reliable, these modes will surely provide much more detailed information about the structure of the magnetic field, just as Brown's observations¹⁴ of nonaxisymmetrical p modes can yield the latitudinal variation of the Sun's internal angular velocity. There is, however, an obvious aspect of the subject where the

analogy does not hold: when Deubner first unambiguously resolved the five-minute oscillations there was a theory of acoustic modes with which to compare the results; for Stenflo and Vogel there is no such theory. □

Douglas Gough is visiting the Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay, India, on leave from the Institute of Astronomy, Madingley Road, Cambridge, UK.

Nucleic acid-binding proteins

More metal-binding fingers

from Jeremy M. Berg

WHEN Miller *et al.*¹ analysed the sequence of TFIIIA, a protein involved in the control of transcription of the 5S RNA gene of *Xenopus*, they identified repeated metal-binding domains (fingers) which they suggested bind to the nucleic acid. Recent reports of the primary structures of several other regulatory proteins, including that of the *Krüppel* segmentation gene of *Drosophila* described by Rosenberg *et al.* on page 336 of this issue², reveal the presence of repeated fingers closely homologous to those of TFIIIA. It seems likely that metal-binding domains are frequently used as structural units in nucleic acid-binding proteins.

Nucleic acid-binding proteins are involved in the positive and negative control of gene transcription. By recognizing specific nucleic acid sequences such proteins play crucial roles in development and gene expression. The determination of the three-dimensional structures of several prokaryotic DNA-binding proteins led to the discovery of a common structural motif used in DNA recognition, consisting of an alpha helix, a short turn and a second alpha helix³. The structurally characterized proteins form dimers such that the two helix-turn-helix units are related by a two-fold axis as shown on the left hand side of the figure. Based largely on the presence of relatively invariant hydrophobic residues that are involved in stabilizing the bihelical structure, similar structures were recognized in several phage and bacterial proteins (see ref.3) as well as certain yeast mating factors and eukaryotic homeo-box proteins^{4,5}. The helix-turn-helix unit became recognized as one frequently used motif for DNA binding.

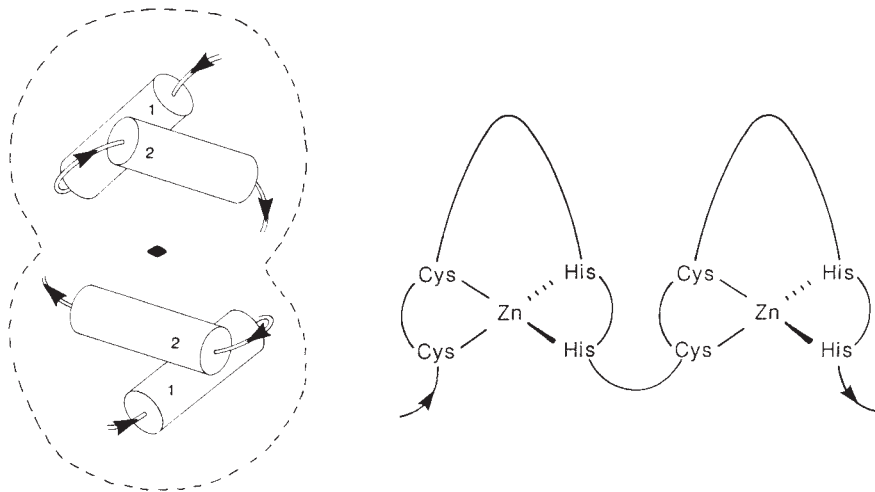
The newly emerging alternative motif of metal-binding domains began with the recognition of nine tandem copies of an imperfectly repeated sequence of about 30 amino acids, including two cysteines and two histidines in essentially conserved positions, with a consensus sequence: (Tyr-Phe)-X-Cys-X₄-Cys-X₃-Phe-X₅-Leu-X₂-His-X₃-His-X₂₋₆, where X is any amino

acid. The underlined Cys and His residues are thought to bind Zn²⁺ ions — the ribonucleoprotein particle containing TFIIIA binds 7–11 Zn²⁺ ions per polypeptide chain — the Tyr, Phe and Leu residues may form a hydrophobic core within each domain, and the domains may bind to nucleic acids by means of hydrophilic amino acid side chains¹. A possible structure is shown on the right hand side of the figure.

Four tandem repeats with a consensus sequence (Phe, Tyr)-X-Cys-X₂-Cys-X₃-Phe-X₅-Leu-X₂-His-X₃-His-X₂, nearly identical to that of TFIIIA, are revealed in the sequence of *Krüppel*² and similar units have also been discovered in the sequence of the *Serendipity* locus of *Drosophila*⁶. Here, the beta protein-coding region has five tandem repeats with sequences very similar to those of the *Krüppel* and TFIIIA repeats immediately preceded by a similar sequence of the form Cys-X₂-Cys-X₃-Phe-X₅-Leu-X₂-His-X₃-Cys-X₂. The homologous delta region sequence has six tandem repeats followed by a non-homologous sequence of about 40 amino acids and then one final repeat. Another example of a protein containing repeated metal-binding domains is ADR1 (ref. 7), a positive regulatory protein from yeast. This protein has two tandem repeats homologous to those from TFIIIA.

A variety of other regulatory and nucleic acid-binding proteins contain sequences which, while not closely homologous to the TFIIIA-like sequences, are nevertheless highly suggestive of metal-binding domains. One example is found in the recently reported sequence of the human glucocorticoid receptor⁸ which has a basis region of about 60 amino acids that contains several Cys-X₂-Cys units and additional Cys residues. Another set of examples is provided by some retroviral nucleic acid-binding proteins^{9,10} that contain either one or two sequences of the form Cys-X₂-Cys-X₄-His-X₄-Cys. Several other proteins contain single copies of sequences with four Cys and/or His residues spaced in a manner suggestive of a metal-binding domain^{7,10}.

1. Stenflo, J.O. & Vogel, M. *Nature* **319**, 285 (1986).
2. Mitchell, J.M. Jr, Stockton, C.W. & Meko, D.M. in *Solar-Terrestrial Influences on Weather and Climate* (eds McCormac, B.M. & Seliga, T.A.) 125 (Reidel, Dordrecht, 1979).
3. Dicke, R.H. *Solar Phys.* **78**, 3 (1982).
4. Isaak, G.R. *Nature* **296**, 130 (1982).
5. Sonnet, C.P. *Nature* **306**, 670 (1983).
6. Eddy, J.A. & Fröhlich, C. *Adv. Space Res.* **4**(8), 121 (1984).
7. Gough, D.O. *Variations of the Solar Constant* (ed. Sofia, S.) 185 (NASA conf. publ. 2191, Washington DC, 1981).
8. Spiegel, E.A. & Weiss, N.O. *Nature* **287**, 616 (1980).
9. Däppen, W. *Astron. Astrophys.* **124**, 11 (1983).
10. Endal, A.S., Sofia, S. & Twigg, L.W. *Astrophys. J.* **290**, 748 (1985).
11. Duvall, T.L. Jr *et al.* *Nature* **310**, 22 (1984).
12. Duvall, T.L. Jr & Harvey, J.W. *Nature* **310**, 19 (1984).
13. Gilman, P.A. *Astrophys. J.* **162**, 1019 (1970).
14. Brown, T.M. *Nature* **317**, 591 (1985).



Left, schematic structure of a dimeric protein containing the helix–turn–helix motif; such proteins bind DNA with the first helix of each unit interacting with the DNA backbone and the second helix lying in the major groove. Right, a possible structure for the repeated domains from a TFIIIA-like protein. Nucleic acids might bind to such a structure by interacting with the end of each domain or, perhaps, by lying in the clefts between adjacent domains.

The discovery of internally-repeated sequences in TFIIIA and of its high Zn^{2+} content led to the proposal of a novel structure based on tandemly repeated domains, each of which is structurally based on a tetrahedrally coordinated Zn^{2+} ion. The conserved hydrophobic residues must play a role in stabilizing the precise polypeptide conformation. One known example of a protein that is built up from small domains is wheat-germ agglutinin¹¹. Its structure consists of four 41 amino-acid domains each of which is stabilized by four disulphide bridges. Given the rapid rate at which proteins are currently characterized, particularly with respect to sequence, more examples of small metal-binding domains can certainly be expected soon.

Many exciting results will follow from studies of the structures, functions and mechanisms of action of these proteins. □

1. Miller, J., McLachlan, A.D. & Klug, A. *EMBO J.* **4**, 1609 (1985).
2. Rosenburg, U.B. *et al. Nature* **319**, 336 (1986).
3. Pabo, C.O. & Sauer, R.T. *A. Rev. Biochem.* **53**, 293 (1984).
4. Shepherd, J.C.W. *et al. Nature* **310**, 70 (1984).
5. Laughon, A. & Scott, M.P. *Nature* **310**, 25 (1984).
6. Vincent, A. *et al. J. molec. Biol.* **186**, 149 (1985).
7. Hartshorne, T.A., Blumberg, H. & Young, E.T. *Nature* (in the press).
8. Weinburger, C. *et al. Nature* **318**, 670 (1985).
9. Copeland, T.D. *et al. FEBS Lett.* **162**, 390 (1983).
10. Berg, J.M. *Science* (in the press).
11. Wright, C.S. *J. molec. Biol.* **111**, 439 (1977).

Jeremy M. Berg is a Jane Coffin Childs Postdoctoral Fellow in the Biophysics Department of the Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

Extinct animals

The mammoths' last migration

from Jared Diamond

MAMMOTH madness reigned in Helsinki last summer when the Soviet Academy of Sciences loaned Finland a unique and comprehensive collection of Russian mammoth memorabilia, including the baby mammoth named Dima and two adults, preserved hair and all in the Siberian permafrost. Thanks to such finds, more is known about the biology of mammoths than about most extinct animals and many extant ones, and the Great Mammoth Exhibit illuminated questions of the distribution, foraging ecology and eventual end of the beasts. The exhibition was, however, the last migration for the Russian mammoths. The Soviet Academy has reluctantly decided not to expose this unique collection to another trip abroad.

Pleistocene mammoths of various species occupied cold tundra habitats of

Eurasia from Britain east to the Pacific and south to the Mediterranean and China, as well as western parts of North America and south into Central America^{1,2}. To judge from the quantities of mammoth bones littering Siberia, the animals must once have been widespread but became extinct about 11,000 years ago¹⁻⁴.

Because of the excellent preservation of mammoth soft tissue in the permafrost, Soviet scientists have been able to reconstruct the diet of mammoths from gallons of frozen stomach contents: grasses, sedges and shoots of willows, birch and alder predominate. How did they reach plants through the snow cover during the long Siberian winter? The shapes of the tusks provide a clue. In young mammoths, the tusks pointed forwards but they became increasingly curved with age until the ter-

minal parts were horizontal and almost perpendicular to the head-to-tail axis. Although ill-suited for defence they would have been well adapted for use as ploughs to clear snow from ground vegetation — mammoths may have wintered in groups in which older members functioned as snow ploughs.

Upper Palaeolithic hunters often killed mammoths and may have been decisive in exterminating them. All North American sites associated with Clovis hunters (the first widespread big-game hunters of the New World, about 11,000 years ago) that contain animal bones also contain mammoth bones; the Escapule mammoth skeleton had two Clovis projectile points among its ribs, the Naco mammoth eight (ref. 5). How were mammoths hunted? Standing beneath the two preserved adult mammoths in the exhibition created gives a vivid impression of the suicidal insanity of attacking such an animal. The question is as difficult to solve as it would be for archaeologists contemplating butchered whale bones in Eskimo middens to deduce whale-hunting techniques without the benefit of present-day observations.

The conventional view, reconstructed in familiar paintings, of near-naked ape-like brutes brandishing boulders and crude clubs at an enraged mammoth cannot be accurate. Men could hardly have survived naked on the Siberian winter tundra, let alone subdued a mammoth with such crude weapons. Nor could they have afforded to run serious risks to themselves in mammoth hunts. Assuming a human band size of 50 with 10 adult male hunters, 10 reproductive-age women, live births spaced at 4 years and one mammoth hunt per month, then no more than 10 per cent of mammoth hunts could have led to a hunter's death, otherwise the death rate would have exceeded the male birth rate and humans would have become extinct rather than mammoths.

Evidently, palaeolithic hunters evolved much more efficient, risk-free methods. Anatomically, they were fully modern and since Clovis hunters specialized in mammoths they were presumably as, or more, skilled than people who now successfully hunt elephants on a casual basis without firearms. Modern Africans and Asians use a great variety of low-risk methods to disable elephants: traps (stockades, pitfalls, footsnare, trunk snares or weighted hanging spears), poisoned arrows, encircling with fire and even spearing or hamstringing after a stealthy approach on foot⁶. These methods exploit detailed knowledge of behaviour of elephants, such as their vulnerability while bathing, fear of fire, use of regular paths and waterholes and reliance on smell to detect danger. Palaeolithic hunters probably exploited similarly detailed knowledge of mammoth behaviour — cave paintings in France and Spain depict traps and pitfalls;

the Lehner and Murray Springs mammoth-kill sites were apparently ambushes at streams. A realistic painting of mammoth hunting should show warmly clad professionals calmly spearing a mammoth already crippled by a trap or ambush.

Did hunting really exterminate mammoths or did they die out solely because of late-Pleistocene climatic changes and retreat of tundra? These alternatives have been much debated, not only for mammoths but also for other large extinct Pleistocene mammals⁴. The accurate radiocarbon dating of very small fossil samples by tandem accelerator mass spectrometry⁷ may provide some of the answers. With this technique, Mead *et al.* have just shown that a species derived from cold climates, Harrington's mountain goat, and a species derived from warm climates, the Shasta ground sloth, both disappeared in Arizona at essentially the same time — $11,160 \pm 125$ and $11,018 \pm 50$ years BP, respectively⁸. This pairing

does not fit a climatic extinction hypothesis but coincides suspiciously with the period of the Clovis hunters. It is also suspicious that the abundant mammoths and many other large New World mammals survived at least 22 Pleistocene glacial cycles, to disappear about the time that Clovis hunters arrived. Using tandem accelerator mass spectroscopy to date more precisely the last mammoths will provide a test of this suspicion. □

1. Agenbroad, L.D. in *Quaternary Extinctions* (eds Martin, P.S. & Klein, R.G.) 113 (University of Arizona Press, Tucson, 1984).
2. Vereshchagin, N.K. & Baryshnikov, G.F. in *Quaternary Extinctions* (eds Martin, P.S. & Klein, R.G.) 438 (University of Arizona Press, Tucson, 1984).
3. Vereshchagin, N.K. *Vestnik zoologii* 3, 32 (1982).
4. Martin, P.S. & Klein, R.G., *Quaternary Extinctions* (University of Arizona Press, Tucson, 1984).
5. Haynes, C.V. *Can. J. Anthropol.* 1, 115 (1980).
6. Johnson, D.L. *et al. Can. J. Anthropol.* 1, 107 (1980).
7. Haynes, C.V. *et al. Archaeol. Eastern North. Am.* 12, 184 (1984).
8. Mead, J. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).

Janet Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

Planetary meteorology

Is there lightning on Venus?

Janet G. Luhmann and Andrew F. Nagy

Is lightning unique to Earth, or does the phenomenon occur on other planets? The answer to this question will have implications for matters ranging from the nature of the atmospheric chemistry and the meteorology of the planets¹ to the existence of life. Because aerosols of volcanic origin enhance lightning production, geological processes must also be involved in this familiar phenomenon².

Venus is particularly fascinating, both because of its proximity and because of its permanent obscuring cloud layers. Because lightning on the Earth is associated with the generation of electrical charges in convective clouds, there seems a good chance that Venus is also a candidate for this type of meteorological activity. But the dynamic and thermal conditions in the venusian atmosphere are probably quite different from those on Earth; the clouds, for example, are composed of sulphuric acid instead of water vapour. If the necessary ingredient of strong convection is absent, and the electrical properties of the cloud constituents or other environmental conditions are not appropriate, venusian clouds should never be brightened by the types of discharges found on Earth. Yet there have been inferences that lightning does occur on Venus.

The Soviet space probes Venera 11 and 12 first detected impulsive very low frequency electric fields below the cloud tops which resemble the lightning-generated 'sferics' seen in terrestrial data³. A more recent reanalysis of Venera 9 spectrometer measurements⁴ found supporting

optical data, although a search with the star tracker on the US spacecraft Pioneer Venus Orbiter had failed to show any evidence for lightning⁵. However, the plasma-wave experiment on that spacecraft, showing a whistler-type electric field noise in the 100-Hz channel in the nightside ionosphere⁶⁻⁸, reinforced the Venera results. (Whistlers are a well known electromagnetic signature of lightning discharges on Earth.)

Scarf and Russell⁹ have extended the interpretation of the electric field observations of lightning by demonstrating that well-ordered magnetic fields with approximately vertical orientation 'duct' the whistlers from the lower nightside atmosphere to the spacecraft located at altitudes above 150 km. Moreover, these ducts, when traced along straight lines in the direction along the magnetic field towards the planet, all appear to intersect the ground near regions identified as volcanic highlands by radar experiments¹⁰. On this basis, Scarf and Russell concluded that the lightning sources on Venus are clustered near specific topographical features where aerosols, such as those associated with active volcanism, may be present.

Taylor and coworkers¹¹ have now suggested that Scarf and Russell's results are simply a coincidence arising from the combination of the sampling bias imposed by the spacecraft orbit, planetary rotation and the magnetic field geometry at Venus. Their main arguments stem from the observation that depletions or troughs in the observed ionospheric ion densities

occur in conjunction with the vertical magnetic field present during the 100-Hz bursts, but that the vertical fields associated with the longer duration troughs do not map preferentially to the highlands. They point out that several other studies show that the vertical magnetic fields are related to the solar wind interaction with the planetary ionosphere causing the interplanetary field to become 'hung up' in the ionosphere^{12,13}, and that it is unlikely that these fields penetrate to the ground. Moreover, the ion-density depletions must also result from the interaction between the solar wind and the ionosphere¹⁴. Similar ion troughs occur in the Earth's ionosphere at mid- and high latitudes in response to magnetosphere-ionosphere coupling. As plasma instabilities can result from gradients in the plasma properties, it is not surprising that plasma waves occur in conjunction with the ionospheric density gradients on Venus. Lightning need not be invoked; therefore the argument for volcanism is moot.

Indeed, evidence continues to accumulate showing that the vertical magnetic fields in the nightside ionosphere of Venus are the result of the solar wind interaction. The ion-density depletions and density gradients are associated with the vertical fields¹¹ but this does not mean that the vertical fields and ionosphere troughs produced by the solar wind interaction cannot duct lightning-induced whistlers to the spacecraft when there is a favourable connection to the highlands.

Because the Pioneer Venus mission has continued to accumulate data, Scarf and coworkers can look for further support for their claims as new areas of venusian topography rotate into the nightside region of vertical magnetic fields. Meanwhile, theoretical investigations of the nature of the alleged plasma instabilities in the ion troughs may help finally to resolve the question. □

1. Rinnert, K. *J. geophys. Res.* 90, 6225 (1985).
2. Veinmeister, P.E. *The Lightning Book* (MIT, Boston, 1961).
3. Ksanfomality, L.V. *et al. Pisma Astron. Zh.* 5, 229 (1979).
4. Krasnopolsky, V.A. *Cosmic Research* 18, 429 (1980).
5. Borucki, W.J. *Icarus* 52, 354 (1982).
6. Taylor, W.W.L., Scarf, F.L., Russell, C.T. & Brace, L.H. *Nature* 279, 614 (1979).
7. Scarf, F.L., Taylor, W.W.L., Russell, C.T. & Brace, L.H. *J. geophys. Res.* 85, 8158 (1980).
8. Ksanfomality, L.V., Scarf, F.L. & Taylor, W.W.L. in *Venus* (ed. Huntten, D.) 565 (Univ. Arizona Press, 1983).
9. Scarf, F.L. & Russell, C.T. *Geophys. Res. Lett.* 10, 1192 (1983).
10. Mazursky, H. *Lunar and Planetary Science XIV* part 2, 466 (LPI/USRA, Houston, 1983).
11. Taylor, H.A., Jr, Grabowsky, J.M. & Cloutier, P.A. *J. geophys. Res.* 90, 7415 (1985).
12. Luhmann, J.G. & Russell, C.T. *Geophys. Res. Lett.* 10, 409 (1983).
13. Marubashi, K. *et al. J. geophys. Res.* 90, 1385 (1985).
14. Brace, L.H., Theis, R.F., Mayr, H.G., Curtis S.A. & Luhmann, J.G., *J. geophys. Res.* 87, 199 (1982).

Janet G. Luhmann is at the Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024; Andrew F. Nagy is in the Department of Atmospheric and Oceanic Science, University of Michigan, Ann Arbor, Michigan 48109, USA.

Cerebral damage and functional localization

SIR—H. Stowell's¹ response to my *News and Views* comment² on the paper by Hart *et al.*³ raises more problems than it solves. I, of course, agree that the brain damage sustained by Hart *et al.*'s patient was indeed global by the criteria of those who work with surgical ablations in animals (not to speak of single cell recordings). But by the standards of human neuropsychology, the infarct was relatively focal; we work with the lesions that nature, sadly, gives us.

Like Stowell, I too have my doubts about direct mappings between categories of cognitive impairment and lesion sites and find no difficulty in providing Stowell with further ammunition for his position. For example: in 1925, Vendryes⁴ wrote that "it is wrong to think of the brain as if it were built on the plan of a grammar, cut into sections for the different parts of speech." Yet in 1985, McCarthy and Warrington⁵ could report a case of severe agrammatism, consequent upon diffuse cortical atrophy, in which a specific disorder of verb use was the most outstanding symptom. If it is any consolation to Stowell, I lie awake at nights worrying about how such phenomena are possible.

Finally, I have the greatest respect for Maxwell, Einstein, Pellionisz, and Llinás, and would much appreciate hearing from Stowell about exactly how we could apply their insights to the interpretation of the highly constrained linguistic deficits seen after (not so focal) brain damage.

JOHN C. MARSHALL

Neuropsychology Unit,
The Radcliffe Infirmary,
Oxford OX2 6HE, UK

1. Stowell, H. *Nature* **318**, 414 (1985).

2. Marshall, J.C. *Nature* **316**, 388 (1985).

3. Hart, J., Berndt, R.S. & Caramazza, A. *Nature* **316**, 439 (1985).

4. Vendryes, J. *Language: A Linguistic Introduction to History* (Routledge and Kegan Paul, London, 1925).

5. McCarthy, R. & Warrington, E.K. *Neuropsychologia* **23**, 709 (1985).

The magnetic field of Uranus

SIR—Reports that the Planetary Radio Astronomy experiment (PRA) on Voyager 2 has not so far detected radio emissions from Uranus, particularly in the previously reported 0.5 MHz region, do not necessarily imply that Uranus has no magnetic field. On the contrary, we think it likely that Uranus will be found to have a polar magnetic field strength of 0.01 to 0.1 gauss, which would imply a magnetosphere extending out to 8 to 16 planetary radii (R_U). But such field strengths would be unlikely to lead to radio emissions detectable by PRA and also imply that the

observed^{2,3} Lyman- α emission from Uranus may only in part be caused by the interaction of the solar wind with the magnetosphere.

The presence of a magnetic field, however small, is suggested by estimates of the internal composition of Uranus leading to the conclusion that the planet has a core constituted of refractory elements (Si, Mg, Fe, Ni, ...) with a radius of $0.3 R_U$ and a mass four times that of the Earth⁴⁻⁷. Such a core would contain about 10% Fe and Ni oxides and, if some of the SiO_2 were in the liquid metal form, there would be a liquid core similar in size to that of the Earth. On this basis, one might assume a molten core with a magnetic moment not very different from that of the Earth, which would in turn imply a field strength at the surface equal to that of the Earth at a distance of R_U , whence the estimate of a polar field strength of 0.01 gauss.

The corresponding distance to the magnetopause on the upstream side, assuming a modified Chapman-Ferraro estimate of the maximum magnetic pressure, is in the range 8–16 R_U , in which case the magnetosphere may extend as far as the orbit of the satellite Umbriel or, at the strong-field end of the range, at least to that of Oberon. A complication, at the present heliocentric phase of Uranus, is that the magnetic axis should be approximately aligned with the solar wind; the funnel, however, should end about two-thirds of the distance to the magnetopause⁸, while the penetration of the solar wind plasma will be determined primarily by the reconnection of the magnetic field lines. Accordingly, we expect that Voyager 2 will enter the magnetosphere at 8–16 R_U , some 2–3 hours before closest approach, that it will miss the dayside funnel or polar cusp and remain in the magnetotail for about 12–24 hours after closest approach.

Using these estimates and the calculations of Siscoe⁹ (modified by a planetary rotation period of 16.3 hours), we conclude that the plasmasphere will extend essentially to the magnetopause with a convection timescale of 14 to 29 hours.

We estimate that the power supplied to the system by the magnetohydrodynamic dynamo driven by the solar wind will be in the range $2-9 \times 10^9$ W, greater than the power supplied by a disk-driven dynamo of the kind considered by Hill and Dessler¹⁰, which, with our parameters, would lie between 1×10^8 and 2×10^9 W. The magnetic field-strength in the tail will range from 1.5 nanotesla near the planet to 0.5 nanotesla at greater distances.

It is also possible to predict the geometry of the auroral emissions on the planet's disk if the magnetosphere is taken to be Earth-like with due allowance for the pole-on orientation of Uranus. On the night side of the planet, hard plasma-sheet

precipitation should produce a crescent-shaped auroral band on one side of the polar cap, while soft polar-cusp precipitation should produce another band on the other side. The same configuration is expected on the day-side on Uranus, but with the difference that the plasma-sheet aurora will be thin and that the polar-cusp aurora will fill a substantial part of the polar cap.

The failure of the Voyager PRA experiment to detect radio emission early in the approach to Uranus is explicable if the planet has a comparatively weak magnetic field. Radio emission will be confined¹¹ to those regions in which the ratio of the electron-plasma and gyro-frequencies is less than unity, which implies that for $B = 0.01$ to 0.1 gauss, there will be no wave generation even at the ionospheric electron densities of about 10^4 cm^{-3} typical of those in the terrestrial ionosphere.

Two other arguments lead to the conclusion that radio emission from the magnetosphere of Uranus may have been too weak to be detected in the early stages of the approach of Voyager 2. The observed radio emissions from the Earth, Jupiter and Saturn, when scaled empirically, suggest that the total power radiated is approximately 5×10^{-6} of the energy flux of the incident solar wind, which implies that if there is any dayside emission from Uranus, the flux would be $1-5 \times 10^6$ W, corresponding to $4-20 \times 10^{-18} \text{ W m}^{-2}$ at a distance of 1 astronomical unit, which is considerably less than the radio power from the three other sources. Further support for the view that the radio emission from Uranus would not have been detected early in the encounter comes from the assumption that it may be permissible to scale the frequency-peak of the emission from a planet with the strength of the magnetic field, in which case the frequency maximum at Uranus should appear at 0.3 to 3.0 kHz, 0.5–5.0 kHz and 5.0–50.0 kHz if scaled from Earth, Jupiter and Saturn, respectively¹². But the lowest frequency channel of the PRA experiment at 0.7–1.7 kHz is below the plasma frequency of the solar wind plasma behind the bow shock, while the next channel (19.9–20.9 kHz) is well above the position of the peak unless the magnetic field at the surface of Uranus is greater than about 0.04 gauss.

For these reasons, we believe that the absence of radio emissions in the early records of the PRA experiment does not imply that Uranus does not have a significant magnetic dipole moment and a magnetosphere of considerable extent.

W.I. AXFORD
V.M. VASYLIUNAS

Max-Planck-Institut für Aeronomie,
D-3411 Katlenburg-Lindau,
FRG

1. Brown, W.L. *Astrophys. J.* **207**, L209 (1976).
2. Durrance, S.T. & Moos, H.W. *Nature* **299**, 426 (1982).
3. Clarke, J.T. *Astrophys. J.* **263**, L105 (1982).
4. Reynolds, R.T. & Summers, A.L. *J. geophys. Res.* **70**, 199 (1965).
5. Hubbard, W.B. & MacFarlane, J.J. *J. geophys. Res.* **85**, 225 (1980).
6. Podolak, M. & Reynolds, R.T. *Icarus* **46**, 40 (1981).
7. Stevenson, D.J. *A. Rev. Earth planet. Sci.* **10**, 257 (1982).
8. Midgley, J.E. & Davis, L. Jr *J. geophys. Res.* **67**, 499 (1962).
9. Siscoe, G.L. *Icarus* **24**, 311 (1975).
10. Hill, T.E. & Dessler, A.J. *Science* **227**, 1466 (1985).
11. Curtis, S.A. *Nature* **318**, 47 (1985).
12. Kaiser, M.L. & Desch, M.D. *Rev. Geophys. Space Phys.* **22**, 373 (1984).

Solar wind interaction with Uranus' atmosphere

SIR—The failure of Voyager 2 to detect radio emission from Uranus even from within 1 astronomical unit may be explained along the lines suggested by Axford and Vasyliunas (preceding letter) by supposing that the surface magnetic field is approximately 0.01 to 0.1 gauss, but it is also possible that Uranus simply has no intrinsic magnetic field. In that case, the solar wind may impinge directly on the neutral upper atmosphere, as it does on Venus. I shall explore briefly the consequences of this assumption for the forthcoming encounter between Voyager 2 and Uranus.

Figure 1 shows the circumstances likely to be encountered by Voyager 2. The scavenging effect of the solar wind should produce a wake of exospheric ions (H_2^+ , He, H_3^+ , CH_4^+ and so on) converging behind the planet^{1,2}. As well as this trail of atmospheric ions, Voyager 2 must be expected to measure a flux of heavy ions sputtered from the particles of which the rings are made. There is thus a chance that some information on the surface composition of these narrow rings may be obtained, while the total area of the optically opaque ring system (approximately $4 \times 10^{17} \text{ cm}^2$, or more than three times the projected area of Uranus itself) may con-

stitute a detectable obstacle to the flow of the solar wind.

Even if Uranus itself has no magnetic field, it is to be expected that there will be a certain degree of amplification of the interplanetary field in the neighbourhood of the planet, as there is in the neighbourhood of Venus; we estimate^{3,4} that a peak field of 4 nanotesla in the sub-solar regions may be induced in this way. Moreover, the interplanetary field lines will be expected to be draped around Uranus, forming a downstream magnetotail, as at Venus and as observed by Voyager 1 during its close encounter with the Saturn satellite Titan.

The intense ultraviolet emission from Uranus must be explained within the same framework. The total energy flux of the solar wind at the projected cross section of Uranus is of the order of $5 \times 10^9 \text{ W}$, which is several orders of magnitude less than the flux required if the observed Lyman- α emission is to be excited directly by charged particle impact. This is the argument by which Hill and Dessler concluded that a mechanism based on auroral excitation would require a sizeable magnetosphere with a surface magnetic field of a few gauss, but if the function of the charged particle impact is to fragment H_2 molecules, which then give rise to ultraviolet emission by the resonant scattering of solar radiation, the energy requirements are less critical. Even so, the upper limit of $5 \times 10^9 \text{ W}$ in the case of a Venus-like interaction is a stringent condition on the energy deliverable to the atmosphere of Uranus by the solar wind.

Support for the view that the Lyman- α flux may arise from processes other than the direct impact of magnetospheric electrons has been provided by the earlier encounters of Voyager spacecraft with Titan and by investigations of the surroundings of Venus, chiefly by means of Soviet spacecraft. One idea is that the effect of the enhanced curvature of the magnetic

field lines around Titan is to increase the flux of suprathermal magnetospheric electrons on the atmosphere of the satellite, which is apparent both in the increased flux of ultraviolet radiation from N_2 molecules and in the depletion of the flux of magnetospheric electrons above 700 eV.

It is important that atmospheric precipitation of charged particles may not be a sufficient explanation. N_2 emission has been observed on the dayside disc of Titan and attributed to the effect of solar radiation on the interaction between the magnetospheric plasma and the atmosphere^{5,6}. Another possibility is that the circumstances around Titan may resemble those around Venus, and also Uranus, in allowing the enhanced ionization of atmospheric constituents (see refs 7–11).

W.-H. IP

Max-Planck-Institute für Aeronomie,
D-3411 Katlenburg-Lindau, FRG

1. Ip, W.-H. in *Future Missions in Solar, Heliospheric and Space Plasma Physics* (Proc. ESA workshop Garmisch-Partenkirchen) 65 (ESA SP-235, 1985).
2. Gringauz, K.I. in *Venus* (eds Hunten, D.M. et al.) 779 (University of Arizona Press, Tucson, 1983).
3. Smith, G.R. et al. *J. geophys. Res.* **87**, 1351 (1982).
4. Neubauer, F.M. et al. in *Saturn* (eds Gehrels, T. & Matthews, M.S.) 760 (University of Arizona Press, Tucson, 1984).
5. Russell, C.T. & Vaisberg, O.L. in *Venus* (eds Hunten, D.M. et al.) 873 (University of Arizona Press, Tucson, 1983).
6. Hartle, R.E. *Adv. Space Res.* **5**, 321 (1984).
7. Alfvén, H. *Rev. mod. Phys.* **37**, 652 (1985).
8. Danielsson, L. *Phys. Fluids* **13**, 2288 (1970).
9. Raadu, M.A. *Astrophys. Space Sci.* **55**, 125 (1978).
10. Formisano, V., Galeev, A.A. & Sagdeev, R.Z. *Planet. Space Sci.* **30**, 491 (1982).
11. Vaisberg, O.L. & Zeleny, L.M. *Icarus* **58**, 412 (1984).

Archaeobacterial terminator

SIR—In his *News and Views* article on "The uniqueness of archaeobacteria" (*Nature* **318**, 234; 1985) Roger A. Garrett quotes me as having suggested "that five consecutive thymidines might be sufficient for ending the methyl-CoM-reductase gene of *Methanococcus voltae*". That is, in fact, not true. Our findings on the terminator structure in *Methanococcus* have since been published (*Nucleic Acids Res.* **13**, 6439–6445; 1985). The terminator strongly resembles a eubacterial terminator.

ALBRECHT KLEIN

Philipps Universität Marburg,
Fachbereich Biologie,
Molekulargenetik,
Karl-von-Frisch-Strasse,
3550 Marburg, FRG

R.A. GARRETT REPLIES—This archaeobacterial terminator was selected from several reported that appeared non-eubacterial in character. Other putative terminators, including those for the ribosomal RNA genes of the extreme halophile *Halococcus morrhuae* and the thermoproteale *Desulfurococcus mobilis*, remain exceptional. □

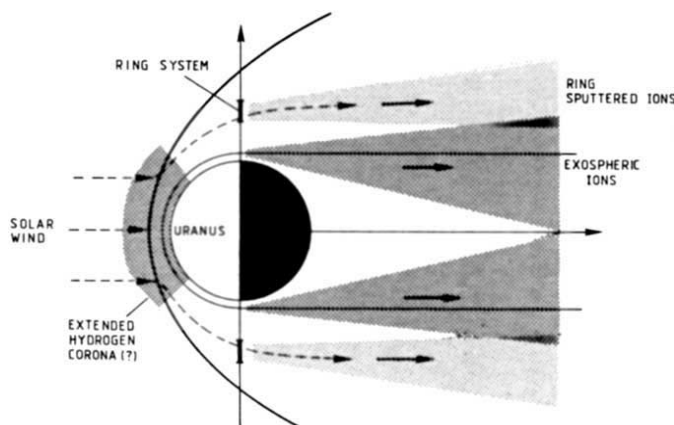


Fig. 1 A schematic view of possible direct solar wind interaction with the neutral atmosphere of Uranus. Points of interest are the upstream ionization interaction with the (putative) extended corona of H atoms; the plasma mantle formed by pickup of the exospheric ions; and the cylindrical sheath of sputtered ions from the rings.

How to test special relativity

SIR—In the light of the recent debate in *Nature* about possible tests of special relativity¹⁻³, we wish to direct attention to a relevant question posed in *Nature* in 1968 by A.C.W.V. Clarke⁴: "What preference is to be given or denied to [the Earth's reference] frame?"

To the best of our knowledge, no subsequent paper in *Nature* has addressed this simple question. We submit that the physical significance of this neglected problem is of a fundamental character, and we briefly explain below how the quest for the answer points to new important tests of special relativity.

The latter theory was built largely on the basis of the following ideas: Owing to the Earth's known orbital motion, the terrestrial laboratory was considered to be a moving frame of reference. Moreover, owing to the continuous change in direction of the Earth's velocity, it was tacitly assumed that the terrestrial laboratory constitutes different reference frames in different seasons (assumption A).

So when the velocity of light was measured (in the terrestrial laboratory), and its value c was discovered to be paradoxically the same in every direction and in every season, Einstein drew his well known conclusions: first, that c may be referred to any and every reference frame; and second, that all frames are therefore equivalent.

If assumption A is granted, this is sound reasoning. Now in kinematics assumption A is quite obviously true indeed. However, despite its apparent plausibility, there is considerable evidence⁵⁻¹⁰ casting doubt on its general validity in many aspects of dynamics, electromagnetics, and optics. In these fields there is some cause to resort to assumption B, that the terrestrial laboratory constitutes at all times one and only one reference frame, namely the unique Earth's frame.

When c is measured in the terrestrial laboratory, the terrestrial environment (atmosphere, gravitational and electromagnetic fields and so on) remains practically unaltered in every season, despite the Earth's orbital and other motions. Because of its size, the Earth exerts autonomy over its domain of influence, and there is no gravitational or electromagnetic field wind, nor a wind of any other kind. So in considerations of this kind, assumption A fails and assumption B seems to be valid.

For these and other reasons, it has been suggested⁹ that the correct interpretation of Maxwell's electromagnetic theory of light might be to postulate that the velocity c of light (Maxwell's electromagnetic waves) has to be referred to the rest frame of the ambient electromagnetic field

(Maxwell's electromagnetic ether) through which the light is propagated, and to this frame only.

In the light of this analysis, all known tests of the special relativistic postulates are rendered inconclusive, for the experiments which led to the inception and testing of the theory of relatively moving frames were all performed in a single stationary frame, the terrestrial laboratory.

Thus from an historical viewpoint, the answer to Clarke's question is that the Earth's frame was unduly given an exclusive preference. From a physical viewpoint, however, the question still remains to be answered. It could not be answered experimentally in 1905 when Einstein published his postulates — he could then perform only *gedanken* experiments.

The advent of the space age, however, has given rise to new possibilities. All sorts of experiments have already been conducted in space. But the few experiments which might have truly tested the perhaps most fundamental and controversial hypotheses in twentieth century physics — Einstein's postulates — have curiously not been done.

The best known, most accurate, and often repeated experiment which led to the genesis of special relativity was first carried out by Michelson and Morley in 1887, and repeated many times thereafter with ever greater precision. Would it not be fitting for a space agency to repeat it for the first time in space during its hundredth anniversary in 1987? M. PSIMOPOULOS

T. THEOCHARIS

Physics Department,
Imperial College,
The Blackett Laboratory,
Prince Consort Road,
London SW7 2BZ, UK

1. Maddox, J. *Nature* **316**, 209 (1985).
2. Tiomno, J. *Nature* **317**, 772 (1985).
3. Aspden, H. *Nature* **318**, 317-318 (1985).
4. Clarke, A.C.W.V. *Nature* **217**, 20 (1968).
5. Essen, L. *Nature* **199**, 684 (1963).
6. Bourdakis, Z.L. *Proc. IREE Australia* **29**, 343-358 (1968).
7. Zapple, C.A. *Speculations Sci. Technol.* **2**, 439-459 (1979); **8**, 61-72 (1985).
8. Aspden, H. *Physics Unified*, 47-69 (Sabberton Publications, Southampton, 1980).
9. Theocharis, T. *Lett. Nuovo Cimento* **36**, 325-332 (1983).
10. Hodges, R. *Electronics Wireless World* **90**, 23-24 (December 1984).

Ontogeny of natural killer cells

SIR—We disagree with the *News and Views* article "Ontogeny of natural killing", in which it was proposed that natural killer (NK) cells are simply immature T lymphocytes, "whose recognition mechanisms never grew up"¹. This is an over-simplification, not substantiated by the present data.

Specifically, we challenge the following statements:

(1) "The (T cell antigen) receptor-positive cells, however, represent only

something like 5 per cent of NK cells: in the rest, although the β -chain genes are rearranged and transcribed², they give rise only to truncated RNA that cannot participate in the formation of a functional receptor." The data in ref. 34 (Reynolds, C.W. *et al.*, *J. exp. Med.* **161**, 1249; 1985) were misquoted. In fact, the title of the article is "Lack of gene rearrangement and mRNA expression of the β -chain of the T cell receptor in spontaneous rat large granular lymphocyte leukemia lines." Furthermore, Ritz *et al.*³ have demonstrated that human CD3- NK clones do not rearrange the β -chain genes. Recently, we isolated CD3-, CD16+ NK cells directly from peripheral blood and have proven that this population does not rearrange the β -chain genes³. Although a truncated β -chain mRNA may be present in CD3- NK cells, B cell lines also produce abortive β -chain transcripts⁴.

(2) It is implied that CD3- NK cells are simply stage I thymocytes, based on the observation that NK cells express CD2, but usually lack CD4 and CD8. By extrapolation, should the reader also conclude that monocytes and a subset of T cells are ontogenetically related, since both express the CD4 antigen? There is no experimental evidence indicating that a CD3- peripheral blood NK cell can be induced to rearrange T-cell antigen receptor genes or differentiate into CD3+ T lymphocytes. Although NK cells and T cells may share a common bone marrow precursor cell, there is no evidence to suggest that NK cells arise from the thymus.

(3) "A careful analysis of T-cell receptor gene expression in parallel with physiological studies on NK cells has now confirmed their membership of the T-cell lineage, but seriously damaged their credibility as a legitimate T-cell subset." Since CD3- NK cells do not rearrange T-cell antigen receptor β -chain genes, and cannot be induced to differentiate into CD3+ cells, what is the basis for concluding that they have been definitively assigned to the T-cell lineage? Moreover, the small CD3+ T cell subpopulation that does mediate non-MHC restricted cytotoxicity, productively rearranges and transcribes the T-cell antigen receptor genes, and recognizes antigen via the CD3/Ti complex⁵. We propose that these non-MHC restricted cytotoxic T lymphocytes do in fact constitute a legitimate T-cell subset, and should be referred to as "non-MHC restricted CTL", rather than NK cells.

Perhaps the underlying cause for this confusion is that there has been no precise definition for an NK cell. Unfortunately, any lymphoid population killing K562 or YAC has been called an "NK cell" or "NK-like" cell. However, the functional, antigenic, and molecular genetic studies on both freshly isolated lymphocytes and

IL-2 dependent clones clearly indicate that both a CD3+ T cell population, and a distinct CD3- NK cell population mediate non-MHC restricted cytotoxicity^{2,3}. There is no evidence to suggest that CD3- NK cells, a population comprising ~ 10% of peripheral blood lymphocytes, are an immature member of the T cell lineage. Hopefully, the relationship of NK cells and T cells will be clarified by the discovery of the NK cell target recognition receptor.

LEWIS L. LANIER

JOSEPH H. PHILLIPS

*Becton Dickinson Immunocytometry Systems,
PO Box 7375, Mountain View,
California 94039, USA*

1. Robertson, M. *Nature* **317**, 768-771 (1985).

2. Ritz, J. *et al. Science* **228**, 1540-1543 (1985).

3. Lanier, L.L., Cwirla, S., Federspiel, N. & Phillips, J.H. *J. exp. Med.* (in the press).

4. Calman, A.F. & Peterlin, B.M. *Fed. Proc.* **44**, 1319 (1985).

MIRANDA ROBERTSON REPLIES—My remark on the possibility that NK cells might be arrested T cells was intended only as speculation: I was explicit about this. Ritz *et al.*¹ did not address the question of rearrangement but state that truncated beta transcripts are made from T3⁺ NK cells; Reynolds *et al.* (my ref. 34) state that beta genes are transcribed at low levels in LGL cells and "the size of the mRNA from the LGL leukaemias is restricted to be 1.0 kb size. A beta chain message of this size is probably transcribed from an unrearranged or simple DJ rearrangement" (*sic*). So it is true that I overstated the data in those two papers: in fact the only claim for rearrangement was by Yanagi *et al.* (*Nature* **314**, 631; 1985; my ref. 33). I also accept that it does not necessarily follow from the data of Siliciano *et al.* (*Nature* **317**, 428; 1985; my ref. 35) that NK cells belong to the T lineage, and my wording was incautious. Finally I heartily agree on the need for a precise definition of an NK cell and the identification of its receptor. □

The search for extraterrestrials

SIR—I welcomed the review article by M.D. Papagiannis on the Search for extraterrestrial intelligence (SETI) (*Nature* **318**, 13; 1985) which, I am sure, gives an accurate view of the global programme.

Faced with the cosmic haystack, I could not help thinking that the approach to SETI so far was somewhere between that of the drunk in the dark who searched for the lost diamond ring under the lamppost, and the rich father-in-law who brought in a bulldozer to do the job properly. The lamppost approach is to some extent inevitable: we have to search in spectral regimes that are technically open to us. But faced with a dauntingly wide spectrum there is opportunity to use a bit more intelligence and a bit less bulldozer when it comes to constructing vast capacity multi-channel spectrum analysers.

First, the microwave window (Fig. 2 of Papagiannis) has a fairly well defined minimum arising from the cosmic background, the galactic non-thermal background and the quantum limit. This will be apparent to any galactic civilization capable of interstellar communications and occurs between 4 and 6 GHz.

Second, the so called magic frequencies of H and OH, while they are most important as absolute standards, are contaminated by natural backgrounds and they do not occur at the optimum frequency in the microwave window.

What SETI needs is a "search filter" which it can apply to the cosmic haystack so as to assign search priorities on the basis of the rational choice of transmission frequency. This choice should combine a physical constraint such as the energy constraint which defines the minimum in the microwave window, an absolute frequency standard such as the H line, and a civilization signature constant (CSC): some absolute number by which the frequency standard can be multiplied or divided to obtain the final transmission frequency.

The most obvious choice for the CSC is π or e , and using the H line we arrive at 4.468 GHz and 3.863 GHz as most likely communication frequencies. Since it is difficult to determine the cultural bias in choosing the CSC, care should be taken in investigating such CSCs as $(e\pi)^4$, π^2 and dimensionless physical constants.

If the systematic searches of the microwave spectrum first concentrate on a narrow frequency band that encompasses the Doppler broadening about the above absolutely defined frequencies, they will increase their chance of early success, without prejudicing the rest of the search.

Unimagined technologies could radically alter the physical constraints. The physical constraint would be quite different for a civilization which is in close proximity to an old and permanently nullified pulsar. In this case the pulsar could very effectively be used as a broad band transmitter of radio pulses, by firing projectiles at it with the appropriate modulation. Since a significant fraction of the projectile rest mass would be converted to radiation, this would provide a powerful means of interstellar communications, with a range of several kiloparsecs. It might be worth examining known pulsars for signs of amplitude or phase modulation.

Papagiannis is right that the answer can only come through "scientific searches . . . pursued vigorously", but we should give "them" credit for some rationality when we devise the search. If they do not want to communicate, and if we are simply eavesdropping, the bulldozer approach may be the only way, but the chance of early success will then probably be limited.

D.G. BLAIR

*Department of Physics,
Nedlands, Western Australia 6009*

Clone controversy at Immunex

SIR—At the Fourth International Lymphokine Workshop in West Germany on 17-21 October 1984, an executive of the company Immunex challenged our claim to have cloned the complementary DNA for interleukin-1 (IL-1) and informed the workshop that we had instead cloned some other monocyte product produced in response to lipopolysaccharide stimulation. Although not willing to present any scientific data to validate his claim at the time, he then added that Immunex had obtained a similar clone to ours several months previously, but had discounted it as being that of IL-1. A full account of this public outburst is now available¹.

We therefore read with great interest the recent article² entitled "Cloning, sequence and expression of two distinct human interleukin-1 complementary DNAs" from Immunex Corporation, of which the same executive is an author. Immunex now obviously feels confident about the fidelity of our human IL-1 cDNA clone since they confirm our work on the molecular cloning of the predominant (pI 7) form of monocyte-derived human interleukin-1 which is now published³.

In addition Immunex has added unnecessary confusion to the field by assigning an α - β nomenclature to the IL-1 sequences. We take issue with this action on two counts. First, we feel that the establishment of such classification is premature when based solely on characterization of cDNA clones. A meaningful classification can only be established when information is available on complexities of genomic organization, post-transcriptional and/or translational processing and the relationship to the multiple forms of IL-1 protein which have been described. Second, in the event that the α - β system were to be adopted, we suggest that it would be more appropriate to call the pI 7 form ' α ' and the pI 5 ' β ' rather than the reverse which was proposed by Immunex. This suggestion is based upon criteria of prevalence of the molecule (50-fold more pI 7 than pI 5) and the chronology of the cloning in the human.

SHELDON M. WOLFF

PHILIP E. AURON

CHARLES A. DINARELLO

LANNY J. ROSENWASSER

*New England Medical Center,
Boston, Massachusetts 02111, USA*

ANDREW C WEBB

*Wellesley College,
Wellesley, Massachusetts 02181, USA*

ALEXANDER RICH
*Massachusetts Institute of Technology,
Cambridge Massachusetts 02139, USA*

1. *Lymphokine Res.* **4**, 51 (1985).

2. March, C.J. *et al. Nature* **315**, 641-647 (1985).

3. Auron, P.E. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 7907-7911 (1984).

Between men and women

Robert Attenborough

The Descent of Woman: A New Edition. By Elaine Morgan.

Souvenir Press: 1985. Pp.288. Hbk £8.95; pbk £5.95.

Myths of Gender: Biological Theories About Women and Men.

By Anne Fausto-Sterling. *Basic Books: 1986. Pp.258. \$18.95.*

THERE are some books which a male reviewer takes on at his peril. These two come into that category. Both authors are female and feminist, and both write about human evolution and biology with a particular interest in the female side (or, to put it more even-handedly, in sex and gender). But there, on the whole, the resemblances end. Elaine Morgan is a Welsh freelance writer, and her speculative evolutionary history, of woman as well as man, dates back to 1972. Anne Fausto-Sterling is an American academic (a developmental geneticist), and her new book surveys a series of controversies about the biology of sex differences and roles.

Elaine Morgan's publishers puff up her book as "the classic study of evolution — a new edition". Certainly it was enjoyed by many, was welcomed as a breath of fresh air by feminists and was found stimulating by at least some evolutionary biologists. If the books of Robert Ardrey, Konrad Lorenz and Desmond Morris (in order of chronology, not merit) represented the first generation of the speculative human ethology which began in the 1960s, Elaine Morgan's belonged to the second. She had read the earlier books, and to some extent the academic literature beyond them, and wrote a riposte in a similar style giving her version of the human evolutionary scenario in the manner that Stephen Gould calls "adaptive storytelling". Her particular theme was a dual one: that evolutionary views of "man the hunter" and so on were too male-centred, not just in using "man" and "he" for the species but in paying too little attention to female roles and selection pressures; and that a better scenario for human evolution could be based on the "aquatic ape" hypothesis — the idea, proposed by Sir Alister Hardy and outlined to a wider public by Desmond Morris, that there was an aquatic (perhaps better, littoral) phase in human evolution during the Pliocene which influenced our adaptations then and since.

Morgan wrote with an engaging panache, and while not all her arguments and criticisms hit their target effectively, some certainly did. That she was right to focus more attention on the female part in human evolution is clear, and other books such as Nancy Tanner's *On Becoming Human* (Cambridge University Press, 1981) and Sarah Hrdy's *The Woman That*

Never Evolved (Harvard University Press, 1981) have since made the same point in different and more scholarly ways. Not so many authors have followed Morgan in the aquatic hypothesis, but she herself wrote a second book, *The Aquatic Ape* (Souvenir Press, 1982), designed, with some success, to win from academics a more serious hearing for the idea. In both its main themes, then, *The Descent of Woman* is now largely superseded. In not revising the book beyond deleting one chapter and adding a 15-page postscript, Elaine Morgan herself seems to admit to a view of the book as more a period piece than an essay to be taken at face value today for its substantive content.

By contrast, *Myths of Gender* certainly does ask, and deserve, to be taken seriously. Rather than tracing an evolutionary sequence or developing a cumulative argument, Anne Fausto-Sterling devotes the core of her book to examining a set of controversial propositions which have in common simply that they attribute some aspect of females, males or the differences between them to biology. Principally, these propositions concern intellectual

differences between the sexes; the development of sex roles in society; the behavioural impact of female hormones in menstruation and menopause; the relationship between male hormones and aggression; and evolutionary (especially sociobiological, needless to say) theories of the behaviour and relationships of the sexes. These are the "myths" of the title, and although that does not strictly imply that they are untrue, the burden of the book is that the propositions in question need sceptical examination, and are often wrong or unsubstantiated.

Fausto-Sterling makes this argument in each case in a way that is in some respects conventionally academic. She herself says that in this politically sensitive area she applies especially high scientific standards, but that is not just an excuse to dismiss unpalatable findings: on the contrary, the book is closely and intelligently argued, well documented factually and carefully referenced (albeit by the irritating end-note method). There are, throughout, interpretations one might argue with, and on occasion there are factual errors (for example that "humans are the only primates exhibiting handedness or hemispheric specialization"). Perhaps, too, the coverage would have been strengthened by greater input from the social sciences.

For whom the book is intended is not entirely evident. Fausto-Sterling can write quite well and clearly, and the book demands little prior knowledge. In some ways, it reads as if designed for a first-year undergraduate audience, perhaps even as

TO CRYSTALLIZATION

The atom is a crystal
of a sort; the lattices
its interlockings form
lend a planarity most pleasing
to the abysses and cliffs, much magnified,
of (for example) salt and tourmaline.
Arise, order,
out of necessity!
Mock, you crystals,
with all appearance of chiselled design,
our hope of a Grand Artificer.
The graceful layered frost-ferns the midnight elves
left on the Shillington windowpanes
for my morning astonishment were misinformation,
as is
the glittering explosion of tinted quartz
discovered in earth like a heart of thought,
buried evidence
crying out for release to the workman's pick,
tangled hexagonal hair of an angel interred
where it fell, our earth still molten, in the Fall.
When, on those anvils at the center of stars
and those even more furious anvils
of the exploding supernovae,
the heavy elements were beaten together
to the atomic number of 94
and the crystalline metals with their easily lost

valence electrons arose,
their malleability and conductivity
made Assyrian goldsmithing possible,
and most of New York City.

Stendhal thought that love
should be likened to a bare branch crystallized
by a winter in the depths of the salt mines of Hallein:
"the tiniest twigs, no bigger
than a tomtit's claws, are spangled with an infinite
number of shimmering, glistening crystals."
Our mathematics and hope of Heaven
alike look to crystals;
their arising, the mounting
of molecules one upon the other, suggests
that inner freezing whereby inchoate
innocence compresses a phrase of art.
Music rises in its fixed lattices
and its cries of aspiration chill our veins
with snowflakes of blood;
the mind grapples up an inflexible relation
and the stiff spheres chime—
themselves, the ancients thought, all crystal.
In this seethe of hot muck there is *something else*:
the ribs of an old dory emerge from the sand,
the words set their bevelled bite on the page,
the loved one's pale iris flares in silent assent,
the electrons leap, leaving positive ions
as the fish-scales of moonlight show us water's perfect dance.
Steno's Law, crystallography's first:
the form of crystal admits no angle but its own.

(Reproduced from *Facing Nature*, a collection of poems by John Updike published today in Britain by André Deutsch, £9.95. Publisher in the United States is Alfred A. Knopf, price \$13.95.)

if the topics were selected more to spark an immediate response in that audience than out of the author's own intellectual interest in them. The way topics are presented, too, suggests an anxiety to keep courting the reader's flagging attention: though the academic content is substantial, it is larded at frequent intervals with turns of phrase that are highly coloured or seek a (contrived) intimacy with the reader — "like the merits of Mom and apple pie . . .", "hormonal hurricanes", "math and sex is a hot topic" and so on. If it is a university audience that Fausto-Sterling has in mind, this surely implies a depressing (but accurate?) estimation of students' stamina for unalloyed academic argument.

What is clear about the intended audience is that it is American. That emerges from the style, the references to "this country" (the United States) and the overwhelmingly American literature that Fausto-Sterling draws on. Indeed, in one sense her book is more revealing about the state of contemporary American society than about the fundamental biology of the sexes, especially in focusing critical attention on some strongly hereditarian and perhaps even social-Darwinist (if that misnomer may be excused) strands surviving in American political life. Fausto-Sterling is greatly and understandably concerned about the presentation of biological theories in the media, and about the tenor of public debates on their social implications, for example on whether arguments about the biological basis of female or male actions should be accepted in the courts as diminishing responsibility for them. This results in frequent references to what *Time*, *Newsweek*, *Mainliner* (apparently an American in-flight magazine) and many others have to say. Up to a point, this is right and proper, since a socially responsible scientist should undoubtedly take notice of what the public at large is reading about science: but it seems to me that some overly superficial treatments are here allowed to determine the agenda and set the tone of the debate.

In short, Elaine Morgan is the more natural and more successful writer of the two, and what she has to say commands attention if not conviction. Stimulating though it was at first publication, however, *The Descent of Woman* is outdated. Anne Fausto-Sterling's *Myths of Gender* is substantial and useful, and has considerable intellectual merits: but, for my money, the author (not uniquely) strives too hard for commercial appeal, and might have been more successful had she had more confidence in the intrinsic interest of her material. □

Robert Attenborough is a Lecturer in Biological Anthropology in the Department of Prehistory and Anthropology, A.D. Hope Building, Australian National University, GPO Box 4, Canberra, ACT 2601, Australia.

At the information exchange

Eugene Garfield, Henry Small and George Vladutz

Scientific Communications and Informatics. By A. I. Mikhailov, A. I. Chernyi and R. S. Giliarevskii. Translated by Robert H. Burger. *Information Resources Press, Suite 700, 1700 North Moore Street, Arlington, Virginia 22209, USA: 1984. Pp.402. \$55.50.*

THE East Europeans have often excelled as theoreticians out of necessity. Lacking a technological infrastructure that allows them the luxury of developing practical solutions to many problems, they lag in applied science but have the time to view developments from the outside. While there are many good theoreticians in the Western world, the distractions of applied research are frequent and often compelling.

The re-publication in English of a leading Soviet book, about eight years after its appearance in Russian, speaks of some strange priorities in the Soviet Union. There the scientist is deprived of the immediacy and convenience of personal access to foreign journals and books; nor are the authorities willing to truly join in international communication by allowing their own journals and books to be published first in English — the lingua franca of international science. Without this misplaced linguistic pride, the seminal research of such as the biochemist Shemyakin would have been known to the rest of the world much earlier. Instead, dozens of notable works by Soviet scientists are doomed to delayed recognition.

How does one properly appraise the translation of a book that itself was not very current? Ten years ago we would have had to say it is a very important work, remarkable as a thorough compilation and analysis of materials on information science (or informatics, as Mikhailov named it many years ago). Today it is still interesting for some of its statistical tables, but there is little in it that is new to the serious student of the subject.

One of the book's most appealing features (although others may find it annoying) is the frequent allusion to historical, sociological and philosophical work which serves to place the field of information science historically and intellectually in the social sciences. Much of what the authors say on the topic of contrasting national styles in information science remains valid, showing how little the field has progressed on the international front. Also, even now the succinct discussion of refereeing — the subject of a vast literature — makes interesting reading.

Here the authors quote John Ziman; it will amuse him to learn that he is an American.

One wonders if the book review editor of *Nature* wasn't a bit mischievous in asking staff of the Institute for Scientific Information to review a work by leaders of its Soviet counterpart. There is, for instance, only a cursory remark (p. 159) about ISI's *Science Citation Index*, in connection with Derek de Solla Price's "half-life" index for literature obsolescence. Price leaned heavily on the *Index* for support for his theoretical notions — as, incidentally, did V. V. Nalimov, who coined the term "scientometrics" ("naukometriia") for the "science of science". But although the ageing effect is illustrated by the interesting discussion on pp. 152–153 of "low-calibre" articles, and Price is quoted, the term "uncited" has been ignored. Similarly, had the book been translated by an information scientist, the term "substantive" might have been replaced by "citation classic". The translator, the authors and Norbert Wiener, whom they quote on the cumulative effects of scientific work, all seem to have been unaware of Ortega y Gasset or the so-called Ortega hypothesis.

More generally, the authors' claim that our failure in the West to read Russian puts us at a disadvantage. This is no more significant than our inability to read Japanese or Chinese; the importance of each "exotic" language varies from field to field and from time to time. The unfortunate effect of the language barrier in the past was in delayed recognition. In the case of Shemyakin we could only have known about his work a year earlier, because it was abstracted. But its publication in Russian, rather than in English, was the principal factor in its relatively low citation compared to the work of David E. Metzler.

There can be little doubt that this book should be in every school of library or information science, though the publisher would have been wiser to urge (or allow) the authors to provide an up-to-date bibliography and a decent index, especially to cited authors. Writing a completely revised second edition will be a considerable task, but no doubt it will occur. We must hope that when it does, the English version will appear simultaneously with the Russian. The publication delay for the translation of the present book, which is by no means uncommon, makes communicating with our Soviet colleagues much like communicating with Mars. □

Eugene Garfield, Henry Small and George Vladutz are at the Institute for Scientific Information, 3501 Market Street, Philadelphia, Pennsylvania 19104, USA.

• Springer-Verlag has published Part I in Series A (original scientific publications) of *Heisenberg: Collected Works*. Price is DM 180. Part IB was reviewed by Nicholas Kemmer in *Nature* 313, 826 (1985).

The bacteria with a difference

D.B. Nedwell

Archaeobacteria. The Bacteria: A Treatise on Structure and Function, Vol. VIII. Edited by Carl R. Woese and Ralph S. Wolfe. Academic: 1985. Pp.582. \$85, £85.

THE aim of this series of books, of which this volume is the eighth, is a comprehensive review of the structure and function of bacteria. Previous volumes have been concerned with structure, metabolism, biosynthesis, etc., but this latest volume is rather different in not being orientated towards a topic but dealing in depth with a specific group of microorganisms.

Archaeobacteria are the group considered, despite accumulating evidence that they are quite distinct from most prokaryotic bacteria. The editors' introduction emphasizes this point, and that the degree of difference between prokaryotic bacteria and archaeobacteria justifies their placement in distinct kingdoms. The inclusion of a volume on the archaeobacteria in a series on bacteria is, therefore, a contradiction, but nonetheless justified as they have in the past been considered as prokaryotic bacteria, which they superficially resemble. Comparison of information in this volume with that on prokaryotes in preceding volumes indicates the basis upon which this fundamental division has been established.

The book is divided into three sections: biochemical diversity and ecology; RNA translation apparatus; and general molecular characteristics. Chapters in the first section deal specifically with each of the major groups of archaeobacteria: the methanogens and extreme halophiles, together with the thermoacidophiles (sulphur-dependent archaeobacteria). Each chapter succinctly reviews the taxonomy, ecology and biochemistry of the group, as far as it is known, and an interesting fourth chapter in this section deals with the geochemical evidence of archaeobacterial evolution from the sedimentary record, including abundance and distribution of archaeobacterial biomarkers.

The second section has four chapters dealing with archaeobacterial ribosomes, nucleic acids and elongation factors: while in the third section, five chapters cover, amongst other topics, cell envelopes, lipids, antibiotic sensitivity and genome structure. This last section tends to be rather a catch-all without the cohesion found between chapters in the first two sections, but this is probably inevitable given the large gaps in current knowledge of the group.

The book is well produced. The standard of writing is uniformly lucid and com-

prehensive, and the text is well illustrated with photographic plates at relevant points. The high standard set by the preceding volumes in this series has been maintained. It pulls together and presents the overall evidence for the distinct nature of archaeobacteria, which should be extremely useful for teachers having to adjust their undergraduate lectures to reflect the many recent developments in the knowledge about these organisms. At its price I doubt whether it will find many private purchasers, but it is predominantly a reference text which, in institutional libraries, will be particularly useful at the postgraduate level. □

D.B. Nedwell is a Lecturer in the Department of Biology, University of Essex, Colchester CO4 3SQ, UK.

Living in the past

Paul G. Bahn

On the Track of Ice Age Mammals. By Antony J. Sutcliffe. British Museum (Natural History): 1985. Pp.224. £12.95. To be published in the United States later this year by Harvard University Press, \$25.

ANTONY Sutcliffe has produced a book which is, presumably, aimed at the kind of interested laymen who visit the Natural History Museum in London. Scholars in the several disciplines on which he draws will merely pick up some new information here and there, but the general public will find a well-produced guide to many aspects of the Ice Age — easy to read, wide-ranging and profusely illustrated (the five colour reconstructions of fauna in landscapes are particularly fine).

The author stresses that he has not written the book as a comprehensive treatise on Ice Age fauna. Instead he has wisely concentrated on a selection of subjects, sites and parts of the world. Inevitably,

perhaps, there is an imbalance caused by emphasis on areas with which Sutcliffe himself is most familiar. Thus readers are presented with rather more information on British quaternary deposits and terraces than they might welcome; the East African evidence and the work of the Leakeys is covered in far more detail than South African material; and while early man looms large in the African chapter, he is barely mentioned in chapters on other parts of the world. This is regrettable since the thorny question of when man entered the New World and Australia has much relevance to the problem of animal extinctions, covered in the final chapter.

The chapter on dating explains some methods and their applications with great clarity, but one wonders why terms such as "isotope" and "amino acid racemization" are left unexplained. Part of the author's task was to show how a specialist in one field needs to draw on researches in others. He has generally succeeded in this, but the difficulty of coping with such a wide range of topics shows through at times, especially in the chapter on palaeolithic art which contains several elementary mistakes: there is no salmon engraving in the floor of Bédouilh cave (it is in Niaux); Kesslerloch cave is in Switzerland; and there was never a continuous barrier of ice from the Atlantic to the Mediterranean along the Pyrenees — Spain was never cut off completely from France.

The fact remains that it is notoriously difficult to write a serious but accessible account of any part of science. Future visitors to the Natural History Museum, and like-minded people elsewhere, should be grateful to the author for providing them with such an elegant and informative book. □

Paul G. Bahn is currently a J. Paul Getty Post-doctoral Fellow in the History of Art and the Humanities, and can be contacted c/o the School of Archaeology, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Frozen find: the carcass of a 6–7-month-old woolly mammoth, discovered near the River Kirgilyakh, Siberia, in 1977. The animal died some 40,000 years ago, and was 115 cm long, 104 cm high and had a trunk 57 cm in length. In life it had a hairy coat, some of which is still visible around the feet. (Reproduced from On the Track of Ice Age Mammals, courtesy USSR Academy of Sciences.)

Genius in a bottle

Augustine Brannigan

Changing Order: Replication and Induction in Scientific Practice. By H.M. Collins. Sage: 1985. Pp.187. Hbk £20, \$25; pbk £9.95, \$12.50.

COLLINS argues that facts are like model ships in glass bottles: the ships are bits of knowledge and the bottles are truth, and once the ships have been assembled, they look like they've always been there and could never come out again. His task has been to study how shipbuilders — the scientists — construct and deconstruct what goes in and out of science. He focuses on the reconstruction of the TEA laser, the gravity waves controversy and disputes over the claims of paranormal phenomena. The work is based on interviews and observations of researchers in each of these areas over the past decade or so, and most of it has already appeared elsewhere. This volume highlights the similar patterns of reality construction and the way conceptual order has been achieved in these areas.

Like most British social studies of science, Collins's approach is brazenly relativist. Order is achieved not because nature dictates what theories work and which do

not, but because the language, models and concepts of science shape how we investigate and experience the world. While this is not new, Collins's work is attractive because he focuses on one of the cornerstones of objectivism: the actual process of replication in science. This is crucial since it is argued that the replication of an effect recommends the independence, the objectivity of the finding. For Collins, what is intriguing is what *counts* as a replication (or falsification) can be very controversial. Generally, unless one is attempting to disconfirm a hypothesis, there is little incentive in exact or identical replication. Also, since publications never spell out (and arguably cannot spell out) all the contingencies in experimental research, replication failures are open to challenge for not being identical. This raises the experimenter's regress:

What the correct outcome is depends upon whether there are gravity waves hitting the earth in detectable fluxes. To find this out we must build a good gravity wave detector. . . . But we don't know that we have built a good detector until we have tried it and obtained the correct outcome! But we don't know what the correct outcome is until . . . and so on *ad infinitum* [p.84].

The most compelling chapter describes the replication of the TEA laser by Robert Harrison (then at the University of Bath) with, we gather, the amateur assistance of Collins. It demonstrates that the replication did not follow an algorithm of inference (nor does induction generally); that the expertise acquired in creating something the first time through is largely tacit and marked by trial and error; that knowledge transfer in publications is quite incomplete; and, finally, that when the procedure works or the model crystallizes the previous sequence appears immediately to be forgotten and the account of how the discovery was made is reconstructed as if it followed an algorithm or recipe, with the experience of contingency and trial and error written off as "human" error. Induction becomes idealized like the ship in the bottle. Science education fosters this picture by directing attention exclusively to completed ships, and ignoring the first-time-through experience.

Change occurs in the relativist world by the construction of novel conceptual systems. These must have an empirical basis, and a conceptual plausibility, yet for Collins what could count as proof is not dictated by a direct grasp of nature, but the negotiation of claims in the institutional or conventional framework. "What counts" is mediated by preferences for different methodologies, different procedures for data analysis and the reaction of the "core set", the cluster of relevant experts. The credibility of a candidate discovery is negotiated through various replications and the scientific community's use of the innovation. Nature does not limit unambiguously what counts as fact, for the

relevance of empirical observation is already imbued with constructs from previous theories. Even the checks on experimenter regress are problematic. For example, the calibration of the sensitivity of gravity wave detectors by introducing electrostatic stimuli to measure their responsiveness entails *a priori* assumptions about the properties of the waves which equate them with other forms of electromagnetic radiation, which, for Collins, puts "constraints on [Weber's] freedom to interpret results" (p.105). However, I think it must be conceded that the acceptance of some such conventions provides positively for the accreditation of evidence, even for, and perhaps especially for, the relativist. The suggestion that gravity waves could have survived beyond their demise in 1975 by use of *ad hoc* hypothesis, and that this would be consistent with the spirit of science, while the destruction of them by failures to replicate was somehow motivated by non-scientific intrigues is not a proposition that is likely to receive much sympathy.

The book gives a vivid sense of the contingent nature of research and is generally a good read, though Collins frequently has a penchant for irreverence and seems to be writing for an introductory audience. My main misgiving surrounds his view of the extra-scientific factors which influence change in science:

For scientific culture, the mediating role of the core set, its laundering of 'illegitimate social interest', and its transubstantiation of social contingency into methodological propriety, along with its privacy, explain the paradox of reification [p.145].

In my view, the core set of experts is unfairly cast as a sort of irrational masonic lodge. No particularly strong evidence is advanced to support this view (though I fear we are into the experimenter's regress?) Also, the idea that illegitimate factors influence the adoption of innovations is not established by discussion of the attribution of character, competence and credibility in the weighing of evidence.

Finally, it seems odd for a relativist to complain that some fact is a reification. This appears to betray a subterranean objectivism, for whatever else could facts be for relativists but reifications? Though the first five chapters are at times brilliant, the final chapter, which deals with the theoretical crux of the matter, how change is accomplished, does not quite get all the pieces into the bottle. □

Augustine Brannigan is Associate Professor in the Department of Sociology, University of Calgary, Alberta T2N 1N4, Canada.

New in paperback

Time's Arrows: Scientific Attitudes Toward Time, by Richard Morris. Publisher is Touchstone, New York (an imprint of Simon and Schuster), price is \$8.95. The book was reviewed by David Park in *Nature* 314, 687 (1985).

nature is
available in microform
from University
Microfilms
International.



☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone () _____

Call toll-free 800-521-3044. In Michigan, Alaska and Hawaii call collect 313-761-4700. Or mail inquiry to: University Microfilms International, 300 North Zeeb Road, Ann Arbor, MI 48106.

Uncertainty and the Strategic Defense Initiative

from John Wright

The latest developments in strategic thinking in the US administration make it necessary to restate classical nuclear strategy.

THERE comes a stage in any debate when it helps to restate basic principles. For nuclear weapons, these were arrived at when the need for some consensus first emerged in the late 1950s and early 1960s. It is because the fundamental facts have not changed but are in danger of being overlaid or overlooked that some restatement is needed now, when we see the possibility of some further agreement emerging between the United States and the Soviet Union and when the view is being put forward that the classic theories are outmoded by new technical possibilities, of which the most important is the 'Star Wars' anti-ballistic missile Strategic Defense Initiative (SDI). The debate itself, both now and twenty-five years ago, often strikes the participant and the observer as both deadly and unreal. The potential for disaster requires the most careful analysis; yet we have so little to work on apart from basic beliefs and aspirations about human behaviour, plus an all-too-detailed knowledge of the capabilities of the weapons themselves. This has always been so and is unlikely to change.

Basic nuclear strategy

In terms of what we can properly call the classical analysis there is little being said now which was not said a generation ago. The nature of that debate is worth remembering. The problem then, as now, was the role of nuclear weapons in achieving some durable strategic military stability. No more was expected or hoped for. The bulk of the work was undertaken in the United States, with much of it being conducted in the Rand Corporation. The conclusion came to then was that stability would be achieved by the possession of an invulnerable retaliatory force capable at all times and in all circumstances of reaching vital targets in enemy territory. The technical implications for command and control, penetration of defensive systems and so on, were worked out in some detail. For such a system to be strategically stable it had to present the same characteristics to the Soviet Union as to the United States. Unfortunately for the subsequent history of the deployment of nuclear weapons, the implications of this analysis could not at the time be accepted and acted upon by either the US or Soviet governments.

The realization that stability depended

on at least a measure of mutual understanding with the Soviet government, if not outright collaboration, came at the wrong time politically. The United States had for the first time been put on the defensive technically—and, it was held, militarily—by the 1957 Sputnik satellite. The emotional impact was reinforced by the realization that the continental United States could, for the first time, be vulnerable to attack by thermonuclear warheads carried on a rocket of a size and power that the United States could not hope to match for many years. Only the unique prestige of President Eisenhower prevented the immediate outbreak of an arms race. But the attempt to come to an understanding with the Soviet government was aborted in the May 1960 summit; the fiction of the missile gap became established and became an element in the 1960 presidential election. The Kennedy administration was committed to correcting it by installing a thousand Minuteman missiles, plus the Polaris programme, as well as retaining the land and seaborne bomber forces of Strategic Air Command (SAC) and the US Navy.

It was at this stage that things started to go wrong. The United States did not wish to give up what many of them saw as a still persisting or recoverable strategic superiority. That this was a dubious concept when the opponent possessed an invulnerable second strike capability was discounted. World peace had been assured by the United States' conscientious policies when they held a nuclear monopoly and this remained a desirable objective, albeit with monopoly diluted to superiority.

This might have been how it appeared to the United States, but the Russians were no more disposed to rely on the US conscience for their continued existence than

they are today. Indeed, the Soviet objective throughout seems to be consistent. There is one theme which is never changed: the need to ensure that the United States (or NATO) does not achieve a strategic superiority which would once again require the Soviet Union to rely on United States restraint for her security. Of equal, if not greater importance is the Soviet government's insistence that the United States admit openly that this is indeed the case. Here we enter into the uncertainties of national psychologies. For whatever reason, be it some ingrained insecurity stemming understandably enough from their own history, or some feeling of continuing inferiority faced with western technical prowess, the Russians do not content themselves with basing their security on the objective reality itself but require this reality to be acknowledged by all.

Faced with the original US monopoly of nuclear weapons, albeit initially in small numbers, they developed their own weapons quickly; they combined this with an army capable of inflicting devastation on western Europe, which to the recipients might not appear all that different from a nuclear strike. The first Soviet missile force developed in strength consisted of 750 medium-range ballistic missiles (MRBMs) capable of reaching all of Germany and most of Europe. The British built up their V-bomber force to about 100 aircraft and the French took the decision under the Fourth Republic to use their existing atomic installations to build weapons.

Although the picture in western Europe must have been satisfactory to the Soviet government—there was no shortage of authoritative spokesmen saying that the Soviet MRBM force could devastate western Europe several times over—this was not the case for them *vis à vis* the

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

The British 'deterrent' in the form of the Avro Vulcan, as seen at the 1957 Farnborough air display. (Photos: Popperfoto.)

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

A South Korean airforce Phantom escorting a Soviet TU-95 ('Bear') in international air space over the Yellow Sea between China and Korea in 1985.

United States, which was beyond their range.

The first attempt to rectify this situation consisted in the construction of a small number of four-engined turbo-prop 'Bear' bombers. They were capable of reaching the continental United States, carrying a thermonuclear device. With this capability Krushchev, for one, believed that he had what we have come rather loosely to call a deterrent. For reasons we shall come to, he was of course quite right. But in the circumstances of the Sputnik and the alleged missile gap, this was an admission which the US government could not make; their public stance was that they were threatened by missiles measured in hundreds, not by subsonic bombers measured in tens. The United States had to catch up. With the announcement of the Minuteman programme of just over one thousand solid-fuelled missiles in hardened silos, the Russians in their turn had to catch up. (In terms of the classical analysis, the possession of a nuclear striking force more powerful than that needed to ensure an invulnerable retaliatory capability did not make strategic sense. Only if one believed that it was possible to eliminate the other side's nuclear forces, thus taking away his second-strike capacity, were there targets to aim at; and those days were over.)

Not long after the Kennedy inauguration, the deployment by the Soviet Union of long range liquid-fuelled intercontinental ballistic missiles (ICBM) in unprotected launch sites had started. Krushchev too thought he was or could soon be facing a missile gap and attempted in Cuba to reproduce in North America some of the characteristics of the situation in western Europe. What happened showed, amongst other things, the limitations in practice of the military effectiveness of nuclear weapons. Krushchev was deterred but the outcome was determined by the local application of a superior US conventional force in the established way.

The Cuban missile crisis brought home the need to establish some understanding between the nuclear powers on the role of nuclear weapons in order to avoid war by miscalculation or by accident. The partial nuclear test ban treaty was signed within a year and a secure 'hot-line' set up between Moscow and Washington. But still there was no formal admission of stalemate or of nuclear parity. Instead we saw the development of McNamara's strategy of 'flexible response'. This permitted the Americans, but obviously not the British or the French, to choose to hit only military targets in a retaliatory or second strike, while still keeping in being sufficient weapons to strike at civilian targets later if need be. This strategy was designed in McNamara's words to give 'a possible opponent the strongest imaginable incentive to refrain from striking our own cities'. In part, this approach was held to be an attempt to answer the question of what happened if deterrence were to fail. This has never been answered satisfactorily and probably never will be.

Unfortunately, and indeed understandably, it is unlikely that the Russians saw the resultant increase in weapon numbers in the same way. In their eyes the force being assembled could, with at least equal plausibility, be construed as an attempt to recreate that strategic superiority which the Soviet government still asserts to be the fundamental objective of US policy. Their reaction was to continue to increase their missile numbers until what they, themselves, would recognise as objective parity had been achieved. But one man's parity is to his opponent a threat of superiority. It is not surprising that the ageing lessons of the classical analysis—especially the sufficiency of an invulnerable second-strike capability for strategic stability—were deemed less relevant.

Successive attempts to achieve some stable and fixed level of nuclear weapons at above the tactical have by and large

failed. How can one equate such technically diverse elements? Still more difficult, how does one secure the political reassurance needed when, in order to do so, one has to take into account such a wholly different mix of past history. Thus, technically the best invulnerable second strike platform is the submarine. To British and to American readers with a history frequently reaffirmed of naval success, this is an idea easy to assimilate and to act on. But to the Russians? A history of naval defeat and a geography of confined access to the open sea has made such a policy difficult to believe in the past, despite the technical arguments. By now of course the Soviet submarine-based missile force is established to a degree which many thought unlikely twenty years ago. Technically, the position becomes more difficult every day. At one stage it might have been possible to reach agreement on a limit on 'throw-weight' or the weight each side's striking force could deliver on to the other's territory. The concept had the advantage of flexibility; numbers of missiles alone did not determine the balance; national preferences could be taken into account, in for example the mix between, say, warheads and penetration aids. We have to face what would seem to be an intractable problem if we are still to attempt to define nuclear strike capacity in discrete numerical terms; some strategic delivery vehicles now have a dual role, with interchangeable nuclear and conventional war-heads. The latest is the Tomahawk submarine-launched cruise missile. Again the problem is not new; aircraft have always had a dual capacity and the attempt to verify the stockpile of aircraft-launchable nuclear weapons has always faced difficulties which have proved insurmountable.

Importance of uncertainty

It is at this stage that another look at one fundamental of the classical analysis is useful. In the discussions now in train we hear much less of an idea which was basic to it: the importance of uncertainty.

To the military, as well as in science, the one thing that can be counted on in the real world is uncertainty. The elder von Moltke told his sovereign that he should be reluctant to embark upon any military operation, even after the most careful calculation, because no one knew what might result. Science is full of uncertainty; the examples are endless. One hesitates also to advance the philosophical objections to certainty in the world of events, though one cannot help feeling that a reading of David Hume would reinforce further the experience of the military and the scientists, should that be necessary.

The fundamental point in the elaboration of any strategy or consistent body of ideas concerning the role of nuclear

weapons is that it is uncertainty which, by inhibiting action, provides us with that stability we seek. In the language of classical nuclear deterrence, we can say that a government is deterred from an action if it cannot be certain that an opponent in possession of nuclear weapons cannot use them to inflict unacceptable damage. The crucial distinction is between those who seek the certain ability to inflict unacceptable damage and those who say that, given the power of the weapons in question, it is sufficient if the government contemplating military action cannot be certain that they cannot be used effectively against it. Returning to Krushchev and his free-fall thermonuclear weapons carried by aircraft, he was right to believe that a United States government, knowing that he controlled such a weapons system, would not use nuclear weapons against the Soviet Union.

What constitutes unacceptable damage has always been a matter of nice debate; it varies with circumstances and by being itself uncertain adds further to the basic stability of the nuclear situation. It is generally accepted that the loss of, say, twenty or fifty major cities in the United States or the Soviet Union in a bilateral exchange would qualify as unacceptable damage in any conceivable set of circumstances; arguably the largest number of nuclear strikes which would make strategic sense was calculated in the Pentagon in Mr McNamara's day to be four hundred of one megaton each. Above this level the incremental increase in effectiveness measured by the number of deaths likely to ensue falls away rapidly. For lesser military powers the objective can be more modest—their ability to destroy one major city in, say, the Soviet Union is likely to constitute unacceptable damage given that the nuclear striking power of the US would remain intact. China has sometimes been held to be a different case; at one stage the leadership apparently believed that alone of the major powers, China could survive an all-out nuclear attack in sufficient numbers to ensure that they won the ensuing war. This view did not long survive the Chinese construction of their own nuclear weapons. It might not have survived an examination of the results of the largest nuclear test ever conducted—the explosion in the atmosphere by the Soviet Union in 1961 of a clean thermonuclear device with a yield in excess of 50 megatons. If the device had been provided with a surrounding layer of natural uranium the yield would have been substantially increased, but more importantly so would the radioactive fallout. Such a weapon could conceivably just make sense if designed for use against a rural population; relatively few would be needed to produce so much fallout as to kill the bulk of the population in the rice-growing areas of China as well as their major cities.

The principle of uncertainty can be extended further. Much effort has been spent on trying to establish what it is that the possession of nuclear weapons with a delivery mechanism capable of reaching an opponent's homeland in fact deters him from doing. Clearly, there is a high probability that it deters him from using nuclear weapons against one's own homeland. But what else? McNamara has said recently that they are 'totally useless—except only to deter one's opponent from using them'. Given the implications of uncertainty, this could be too restrictive a view.

While it is logically possible that, despite its declared policy of doing so, NATO forces might not use nuclear weapons in defence of western Europe against a conventional attack, no certain reliance could be placed upon such an outcome. The posed threat of escalation is effective because no one can risk basing a course of military action on any *expectation* that escalation will not in the event happen; only certainty would do and that is not attainable. The uncertainty is increased when an evident mechanism for escalation is in place and is known to form part of the training of the armed forces stationed in the areas concerned. The Russians appear to share at least part of this view when they say that any conflict in western Europe is bound to become a nuclear conflict, and nuclear conflict is indivisible. Clear statements from the Soviet side are not all that easy to find, especially in the lucid and, we hope, logical form we like to expect in the West. Perhaps Krushchev said it all when he said in 1964: "Nuclear war is stupid, stupid, stupid! If you reach for the push button, you reach for suicide". Nevertheless, in 1969 the Soviet negotiators to the treaty banning unlimited antiballistic missile (ABM) deployment underwrote the principle of deterrence when they said: "Even in the event that one of the sides were the first to be subjected to attack, it would undoubtedly retain the ability to inflict a retaliatory strike of crushing power".

While their daunting power and uncertainty as to whether they might or might not be used makes military leaders pause as never before, they also find themselves faced with a situation on the ground that no responsible military commander can tolerate. Hitherto in war the prudent general has always had in his mind some way of recovering a situation should his well-laid plans go wrong. Nuclear weapons make such prudence untenable; once embarked upon, recovery is impossible. In more general terms we are justified in believing that a proper fear of the unknown future inhibits the possessors of nuclear weapons across a whole range of circumstances in their dealings with opponents who also possess these weapons.

The importance of uncertainty extends

too to defensive systems, in particular to the point-defence systems subject to the control of the 1972 Anti-Ballistic Missile agreement between the USA and the Soviet Union, and now to the more ambitious proposals for the interception and destruction of incoming missiles by a system some of whose components would be stationed in outer space.

Relevance to SDI

The 1972 ABM Treaty and the associated SALT I agreements were made possible when both the US and Soviet governments became convinced that the defensive measures available to them, if deployed, would not alter the strategic balance but would accelerate the arms race. The agreement prohibited all but a limited deployment but permitted research. (The authors of the Treaty understood that research could not be restricted by agreement, the crucial difference here being between research designed to protect against one's opponent stealing a damaging march and research designed to lead to a forbidden deployment. The former used to be US policy in the ABM field; it is in danger with SDI of changing to the latter and, in the course of doing, so generating its own momentum towards deployment).

The issues posed by the SDI are similar in most respects to those posed by the ABM. The debate which led to the 1972 ABM Treaty limiting deployment of defensive systems and to the first agreement to place a limit on some strategic nuclear weapons established that the latter was not feasible without the former, and that any significant deployment of defensive systems would most probably have as its consequence an increase in the opposing side's offensive missiles. As it was then deemed that the cost of providing a quantum of defence was higher than the cost of providing the increase in striking force needed to render it strategically nugatory, it was concluded sensibly enough that everyone would be better off if any defensive deployment were to be strictly limited. At no stage did anyone advance any plausible argument to suggest that anything other than a point, as opposed to an area, defence was feasible. Technically the argument was—and is—that discrimination between incoming objects becomes sufficiently easy to permit the operation of some practical intercepting mechanism only when they re-enter the atmosphere. Outside the atmosphere, the argument goes, plastic balloon decoys behave in much the same way as warheads—and even more like warheads enclosed in plastic balloons. Thus an attack could be attenuated but not eliminated. While this could be held to be of some strategic value in protecting the fixed land-based missiles, it provided no solution to the problem of protecting populations. The one major respect in which the SDI differs from the

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

The US Navy's Tomahawk cruise missile under test at Dugway Proving Ground, Utah.

earlier ABM proposals lies in its emphasis on the attempt to destroy missiles in the boost phase, and thus offer some prospect of area defence. The technical arguments cannot be resolved on the basis of present knowledge but the relevance of uncertainty reasserts itself.

The crucial point in judging the relevance of SDI to the strategic situation we all now face is that any government seeking to base its policy on a defensive system and to discard retaliatory weapons must be certain amongst other things that at all times (including the most difficult—the first time) all incoming nuclear warheads directed at its cities will be destroyed. In an alliance it could be that the same criteria must be made to apply convincingly to its allies also.

Leaving aside the technical difficulties of believing this could ever be the case, we need to recall the possibility of counter-measures. Again we do not know on the basis of present knowledge how effective such measures might be; but they are unlikely all to be completely ineffective. It is calculated, for example, that the burn time for an MX type missile could be reduced from its present 120 seconds to some 60 seconds at a sacrifice of about 18 per cent of payload; there is the possibility of armouring or rotating to protect against beam attack; decoy 'boosters' might be ignited; and of course all equipment in space is easy to find and is vulnerable. Be that as it may, for as long as it is the case that the cost of overcoming a defence remains less than the cost of erecting one—and there is no evidence that the present relationship might be reversed—then the appropriate response to the SDI would not be to construct another SDI, but to increase the power of the offense. This would appear to be the policy of the Soviet government; the only unfamiliar component would be space-stationed weapons designed to destroy the space-borne element of SDI. It is worth recording a Soviet statement made soon after President Reagan's announcement of the SDI. The following is an extract from an interview which Mr Andropov gave to Pravda on 27 March 1983.

President Reagan has announced the beginning of work on a wide-scale and highly efficient anti-missile defence system. At first sight this might sound attractive to the ill-informed, since the President spoke only of

what seemed to be defensive systems. This would, however, be a superficial impression emanating from those unfamiliar with these matters. In fact, the US would continue to develop its offensive strategic forces rapidly and with one end in view, that of acquiring a first strike nuclear capability. In such a situation, the intention of being able to destroy the other side's weapons by means of an anti-missile defence, and thus deprive it of its ability to make a retaliatory strike, is designed to disarm the Soviet Union in the face of an American nuclear threat. (Reprinted in translation in the *Speeches and Writings of Y. V. Andropov*, Pergamon, Oxford, 1983, p. 302).

Mr Gorbachev's views appear to be similar. Indeed immediately after the Geneva summit he made a point of saying that in the event of any likely deployment of space defence systems by the United States, the Soviet response would be effective, cheaper and quicker. He might have had in mind the updating of what many consider to have been the first fully effective Soviet weapons system capable of landing nuclear warheads on the continental United States, namely the placing of two short-range surface-to-surface missiles each armed with a 440 kiloton warhead on a diesel powered submarine capable of reaching the eastern US seaboard. There are other possibilities.

One of the many difficulties in dealing clearly with the arguments for the SDI is that they tend to change according to the proponent and the audience. President Reagan states that the objective of his SDI is indeed to render nuclear weapons impotent and obsolete by providing a perfect defence for the inhabitants of the territories of the United States and its allies. (Nuclear weapons cannot of course be eliminated; both because of the impossibility of accounting for and verifying existing stockpiles and because of the relative ease with which they may be manufactured by any nation with the necessary knowledge and skills at its command.) Others who share the doubts that such an objective is attainable, whether they argue in terms of uncertainty or otherwise, will speak in favour of the system on the grounds that it will produce some useful measure of additional protection for land-based missiles. Others who might well regard the fate of land-based missiles as being largely irrelevant to maintaining strategic stability, tend to be heard less.

Discussion in terms of only land-based missiles can obscure important issues. It permits one to talk of a 'window of vulnerability' which, starting from a consideration of US land-based missiles alone, is applied by extension to the strategic situation of the NATO alliance as a whole; the fact that for evident reasons the majority of NATO strategic weapons are seaborne tends to be overlooked. So too is the still formidable penetration power of a conventional manned bomber force, whether or not enhanced by the supersonic bomber or by aircraft designed to diminish their detection by radar. Neglected too in this equation is the ability of relatively cheap cruise missiles whether land or sea based to penetrate any foreseeable defences, let alone the existence of clandestinely delivered nuclear weapons strategy which even now probably cannot be proved not to be in place. Secretary of Defense Weinberger has put in a logically irrefutable bid for a massive anti-bomber defence system on the grounds that the SDI is effective only against a certain range of weapons. One awaits the bid for an expanded programme designed to offer assured defence against the cruise missile, land or sea based.

Be that as it may, if we believe that a government is deterred as long as it cannot be certain that its opponent's thermonuclear weapons cannot be delivered, it is not possible today to conceive of a situation where mutual nuclear deterrence can fail for technical reasons, and the margins of safety are considerable.

The political arguments

It would be wrong though to leave the analysis at the technical/strategic stage. The point of strategic stability can only be political. Fortunately, political considerations reinforce the stability created by the nature of nuclear weapons and the military immobility they induce. Instead of considering matters such as second strike capability and unacceptable damage and so on we have to face the more fundamental question of whether there can be any circumstances in which a government could believe that it would be better off as a result of initiating the use of nuclear weapons (or indeed other forms of military action) against a nuclear power or its allies. Given the risks involved, which stem from the power of the weapons themselves, it is difficult, indeed probably impossible, to devise such a political set of circumstances. To put the extreme case, what possibility of uncertain political gain could make a government believe it would be better off if, as a result of seeking to achieve that gain, it brought about the destruction of its own largest city? Or indeed even its own probable destruction? In practice no nuclear power would allow its ability to strike to be so limited; not one, but many cities would be at risk. The

political arguments reinforce the military and technical ones to induce that state of paralysis which has helped preserve the peace for the last forty years.

Implications for allies

Although the basis of mutual nuclear deterrence between the United States and the Soviet Union will not be disturbed by the SDI the matter does not end there. The security of other nations is involved also. The debate is only now beginning, but two distinct though complementary points have emerged: the first suggests that any increase in protection that defensive systems could supply would also increase the probability that the United States would risk nuclear war in the nuclear defence of Europe. The second states that if both the US and the USSR were to be in a position to disregard one another's nuclear weapons as they affected their national territories, then this would increase, not diminish, the probability of a 'tactical' nuclear war in Europe, whose effects could hardly be distinguished by its inhabitants from all-out nuclear war. Using Soviet terminology, nuclear war would no longer be indivisible—and hence would be more likely. Europe would, it is argued, become a bull-ring for the superpowers. It is difficult to see how either argument can stand up to precise examination; although the hope has been held out that one day in some way the US, which must be taken to mean US cities, might be made invulnerable to a Soviet attack, this must remain a dubious proposition for the reasons gone into. These arguments too fall down on the impossibility ever of being able to remove the last element of uncertainty in any calculation as to whether the Russian nuclear force could, and in the event

would penetrate US defences, and vice versa. Military planners in Europe and Japan and their political masters are, however, unlikely to see their problems so simply in this light. Any such development would be likely to lead to pressure towards a strengthening of national nuclear forces where they exist and to increasing pressures to create them where they do not. The more the Americans appeared to believe in the effectiveness of their defensive systems and thus tended to withdraw from a commitment to the nuclear defence of Europe, the greater the pressures for more and better national systems in Europe. The implications are incalculable but almost uniformly unpleasant.

Although unrealistic, the declared intention of the US President to render nuclear weapons impotent and obsolete, without, it would seem, thinking through the implications of achieving such a state of affairs, is not reassuring to Europe—strong though its appeal must be to American isolationists. Any move to restore the worlds of 1870, 1914, 1940 and 1941, when state power meant mass armies and where attacking first was held to be a basic rule of military prudence, would be deemed correctly to be reducing European security. It would, by opening the way again to blitzkrieg as a rational military option, mean replacing the present stalemate by one of a conflict in which for reasons of space as well as strength the Soviet Union would have the overwhelming advantage.

Neither are they likely to be reassured by undertakings to consult before deployment when they remember that they were not asked for their views on the proposal before it was first made public; or by any offer to use some of the best of their own

scarce scientific resources on a project which for them is militarily useless and where any other advantage is likely to be slight—if only because of the US record and legislation on the sharing of technology.

Fortunately few of these are much more than debating points; they do not touch the vital issues of strategic stability, which will remain unaffected by SDI. However, unless there are agreements to limit the deployment of new and even more ingenious nuclear delivery systems and their defensive counterparts, we might well have to face up to further substantial expenditures which will nevertheless leave us strategically where we are now. Some might argue that it is this bringing to bear on the Soviet Union of further economic pressure and an increasing demand for scarce skills which forms the true basis of US policy; it is of evident concern to the Soviet government who can ill afford seeing these resources used in this way when there are so many other pressing demands upon them.

Whether that is the case or not, the fundamental facts remain the same now as they were at the time when classical nuclear strategy was first formulated; first, the most important characteristic of a nuclear weapon is its unprecedented power to destroy; and, second, no government can ever be sure that nuclear weapons cannot be exploded on its territory. As a consequence, governments are deterred from a whole range of military actions. We sit on our hands and call it peace—which indeed it is. □

John Wright, formerly on the staff of the Chief Scientific Adviser at the Ministry of Defence, is at 47 Brunswick Gardens, London W8.

REVIEW ARTICLE

Ocean variability in relation to living resources during the 1982–83 El Niño

Richard T. Barber & Francisco P. Chávez

Duke University, Beaufort, North Carolina 28516-9721, USA

Ocean/atmosphere perturbations that generate large-scale thermal anomalies in the Pacific are natural processes that cause significant variability in living resources. Analysis of the 1982–83 El Niño indicates that depression of the nutricline set in motion a decrease in productivity that eventually affected survival and reproduction at higher trophic levels. This impact, together with large-scale southward redistributions, accounted for most of the observed changes in the abundance of living resources.

EL NIÑO is the name given to a series of ocean and atmosphere changes that spread across the Pacific from Australia and Asia to South America, then move to the north and south so that the Americas are affected from Alaska to the tip of Chile^{1,2}. Although El Niño originally referred to a warm surface current that flows southwards along the coast of Ecuador and northern Peru¹, the

term El Niño is now applied to interannual episodes of anomalously warm water, high sea level, depressed thermocline and greatly increased rainfall that extend over most of eastern tropical Pacific Ocean. El Niño is one facet of a global process that is called the El Niño Southern Oscillation (ENSO) phenomenon. Philander¹ and Gill and Rasmusson² have recently

described ENSO phenomenon paying particular attention to the severe 1982–83 El Niño. Here, we shall use their descriptions as a basis for understanding the effects of the 1982–83 anomaly on living resources of the eastern tropical Pacific Ocean. El Niño, through its impact on the global atmosphere and the low- and mid-latitude ocean, is a major natural process causing variability of living resources in the ocean. Some biological and economic consequences of the 1982–83 El Niño are well documented³, but the widespread ecological impacts of this phenomenon are still not fully understood. When considering El Niño it is customary to refer to 'low latitude' effects, but reports of significant ecological changes in the high-latitude Southern Ocean indicate that ecological effects may extend far beyond the tropical ocean^{4–6}.

The coastal ocean off the west coast of South America is the most productive region of the world ocean⁷; this richness depends on coastal upwelling. Coastal upwelling is a circulation pattern that overrides both the nutrient limitation characteristic of warm, permanently-stratified waters, and the light limitation characteristic of deeply-mixed waters⁸. Coastal upwelling is set in motion by equatorwards winds that transport water offshore, resulting in water flowing up close to the coast to replace the water moved offshore. The subsurface waters of the ocean are cool and separated from the surface layer by a thermal gradient, or thermocline, below which the cool water is rich in inorganic plant nutrients such as nitrate, phosphate and silicate. Vertical transport of subsurface water provides continuous injection of new nutrients to the surface and the formation of a shallow wind-driven surface layer, which is vertically stabilized by a thermal gradient, provides optimal light conditions for primary production⁹. Productivity enhancement continues as long as winds are upwelling-favourable, so the annual primary production is much higher in low-latitude upwelling than it is in temperate or high-latitude ecosystems. Abundant phytoplankton are eaten by herbivores and, through grazing and predation, the abundant organic matter passes to zooplankton, fish, benthic animals, birds and marine mammals⁹.

Among ocean phenomena affecting man, coastal upwelling has a relatively direct impact on the social fabric. Traditionally, it is stated⁷ that upwelling regions, while comprising <1% of the ocean's surface, provide more than one-half of the protein collected from the sea. Whatever the proportion, a significant part of the world's fish catch is directly dependent on the physical processes of upwelling. This relationship is illustrated by comparison of the annual temperature anomaly off Peru with the annual catch of anchovy and sardine (Fig. 1). When the ocean is warmer than average the catch is reduced, indicating that warm anomalies such as El Niño are associated with reductions in the abundance of fish and other living resources. Using observations (Fig. 2) from the 1982–83 El Niño, we describe how this reduction takes place and speculate on the speed and nature of the ecosystem recovery.

Environmental setting

Coastal upwelling is a response of the ocean to local winds; in actual dimensions the winds at a specific site affect the ocean for a region as small as 50 km around that point⁹. However, the winds along the west coast of South America are part of the trade-wind system of the Pacific, so variations in coastal winds are related to changes in large-scale winds. Barnett¹⁰ has shown that sea-surface temperature off the west coast of South America is best predicted by variations in the trade winds in the western Pacific, not from wind variations in the eastern Pacific where the upwelling actually takes place. The mechanism responsible for this dependence of temperature on remote winds is that the trade winds set up a basinwide east/west slope in sea level¹¹ and the thermal structure of the Pacific^{12–14}. During normal conditions the persistent easterly trades exert a westward frictional drag on the ocean resulting in the piling up of warm water in the western Pacific¹¹. This dynamic accumulation of mass in

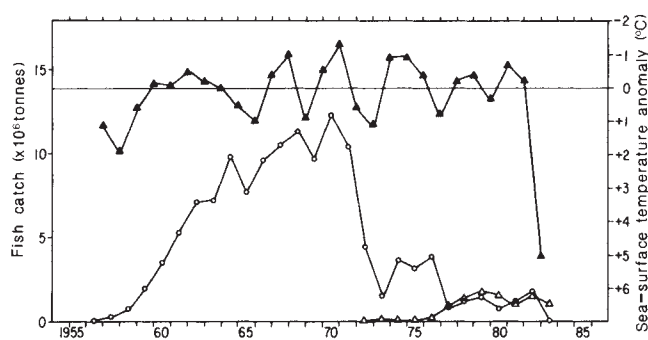


Fig. 1 Physical and biological coupling in the eastern tropical Pacific shown by the relationship between the annual anomaly of surface temperature and the annual catch of anchovy and sardine of Peru. The anomaly temperature scale is inverted: upwards indicates cooler temperatures; downwards indicates warmer temperatures (El Niño). The annual temperature anomaly is relative to the 26-yr mean temperature of 16.9 °C at Chicama. The anomaly is calculated for the Southern Hemisphere thermal year from July to June spanning half of 2 calendar years; the value is plotted in the middle of the thermal year at January. The fish catch for each calendar year is plotted in the middle of the calendar year at July. ▲, Anomaly; ○, anchovy; △, sardine. Updated from ref. 8.

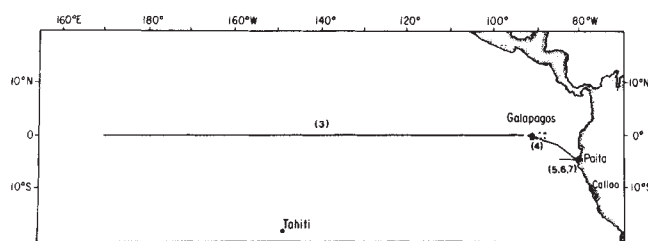


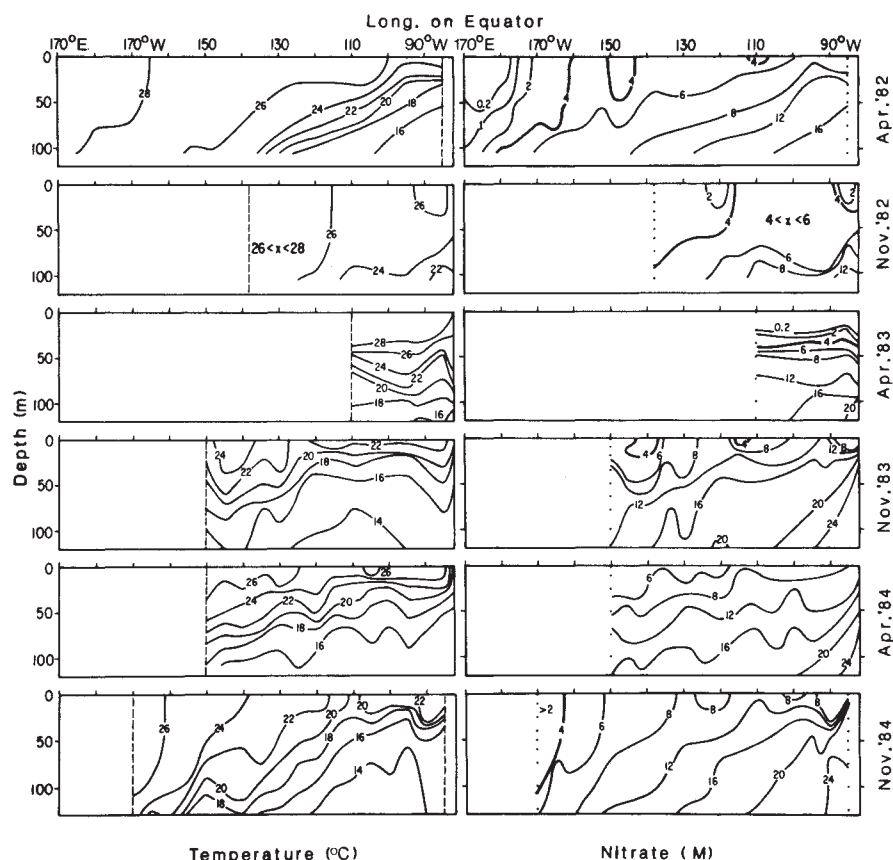
Fig. 2 Chart of the equatorial Pacific Ocean showing the transects and locations that provided our observations. The numbers in parentheses next to the transect or location show the results from that transect or location. Note that the equatorial transect leaves the Equator west of the Galapagos Islands and passes through 2° S and 85° W and ends at the coast of Peru at 5° S and 82° W.

the west produces a sea-level slope of ~0.5 m higher in the west and a thermocline slope of the order of 100 m with the west being deeper¹² (Fig. 3). This condition stores the potential energy to drive a rapid redistribution across the Pacific if the trade winds weaken or reverse.

The east/west slope that exists during normal trade-wind conditions brings the thermocline close to the surface of the ocean off the west coast of South America (Fig. 3). By controlling the depth of the thermocline, large-scale winds control the temperature of the water that feeds into the upwelling circulation because coastal upwelling entrains water from a relatively shallow depth of 40–80 m (ref. 9). The gradient separating nutrient-rich subsurface waters from nutrient-depleted surface waters is called the nutricline. The thermocline and nutricline in the eastern portion of the Pacific (Fig. 3) are much shallower than they are in the central or western Pacific because of the basinwide slope. Thus winds have two relatively separate connections to the coastal upwelling ecosystem: (1) they provide the local driving force for vertical transport; and (2) they set up the basinwide thermocline and nutricline slope that determines the temperature and nutrient content of the water that is entrained into the upwelling circulation. Obviously both local and large-scale winds of the Pacific must be favourable to provide the climatological mean conditions of high nutrient levels and high rates of biological productivity.

Analyses of coastal wind measurements in the past 5 yr have indicated that in previous El Niño events¹⁵ and in the 1982–83

Fig. 3 Nitrate and temperature profiles on the equatorial transect (Fig. 1) showing changes in thermocline and nutricline tilt before, during and after the 1982–83 El Niño. Nitrate is presented as an indicator of the nutrient condition; parallel observations are available for phosphate, silicate and nitrite. Updated from ref. 14.



El Niño^{16,17}, the equatorwards winds driving coastal upwelling did not weaken. During the 1982–83 El Niño, coastal winds and locally forced upwelling appeared to continue through March 1983, but after November 1982 the surface waters had significantly reduced nutrients and increased temperature (Fig. 4). The explanation of these changes that occurred during local upwelling was that the thermocline and nutricline were progressively depressed below the depth where source water was entrained into the upwelling circulation (40–80 m). Figure 3 shows the large-scale thermocline and nutricline topography for April and November of 1982, 1983 and 1984. In November 1982, the thermocline and nutricline were clearly depressed and by April 1983 the characteristic east/west slope had gone¹⁸. Figure 4 shows day-by-day changes in temperature and nitrate at an ocean station in the Galapagos Islands (at ~ long. 90° W in Fig. 4) in late 1982 and 1983. The elevated ocean temperatures that delineate El Niño are clearly associated with a reduction in surface nitrate. The surprising conclusion from these observations is that nutricline depression, not cessation of upwelling, is the process that initially reduces the nutrient supply to the surface layer. This reduction caused primary production to decrease and eventually affected the higher trophic levels of the equatorial and coastal upwelling ecosystems. Upwelling ecosystems are therefore vulnerable to remote ocean/atmosphere perturbations such as the trade-wind anomalies that characterize El Niño Southern Oscillation phenomena^{1–3}.

Progression of the 1982–83 event

The 1982–83 El Niño reached the Galapagos Islands in August 1982 (ref. 19) and the coast of Peru at 5° S in late September 1982 (ref. 20). Figure 4 shows the progression of changes in temperature and nitrate at the surface and at 60-m at an ocean station in the Galapagos Islands. The transition phases of onset in September 1982 and recovery in June 1983 are best delineated in the temperature changes at 60-m. The variation in nitrate

(Fig. 4) shows that El Niño enormously reduced the surface nutrient concentration. At the beginning and end of the time series shown in Fig. 4, surface nitrate values were 4–6 μM . The large-scale zonal profiles of nitrate shown in Fig. 3 show that this concentration range is the mean condition for the region during normal or non-El Niño conditions.

The relationship between phytoplankton growth and ambient nutrient concentration is complex^{21,22}, but for understanding the ecological impact of El Niño it is adequate to know that

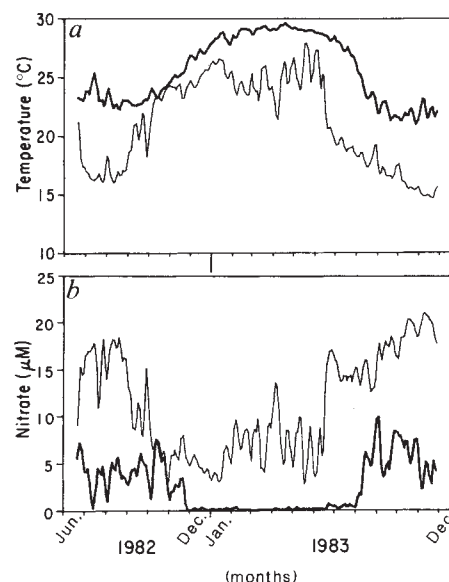


Fig. 4 Time series of triweekly observations of *a*, temperature and *b*, nitrate made during the 1982–83 El Niño on Santa Cruz Island in the Galapagos Islands. Bold line, 0 m; lighter line, 60 m.

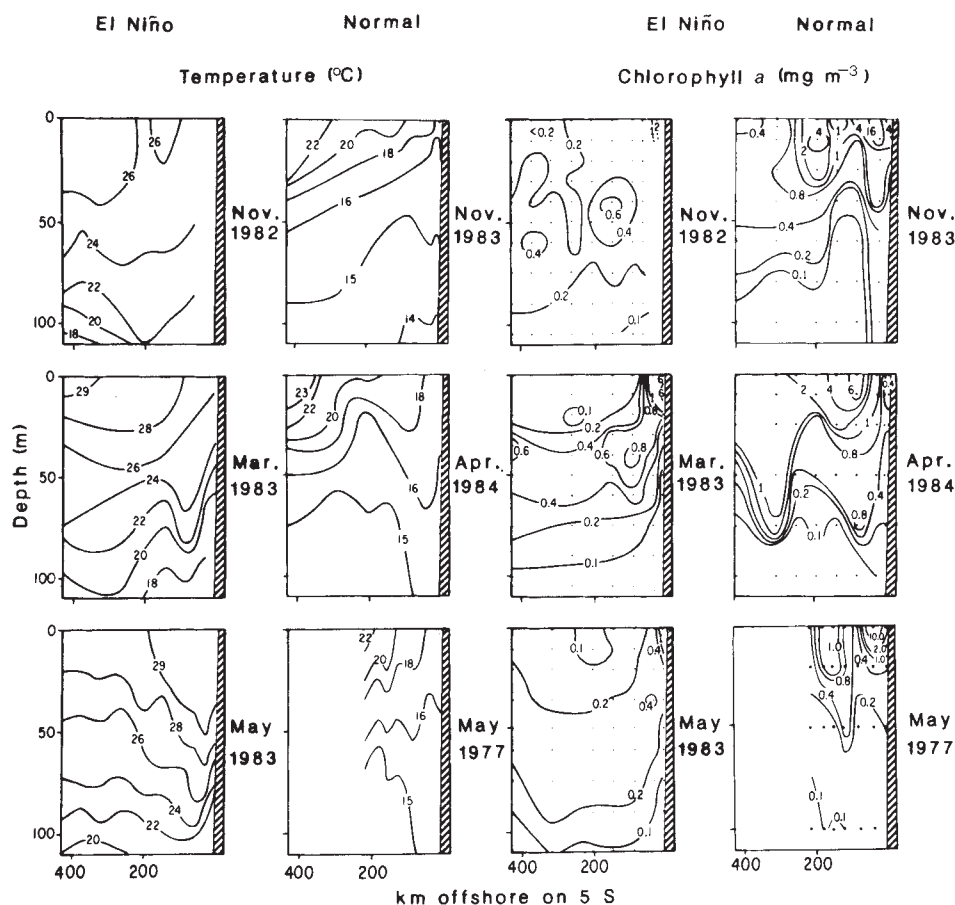


Fig. 5 Cross-shelf profiles of temperature and chlorophyll along a transect at 5° S from 85° W to the coast of Peru. The sequence shows conditions during onset (November 1982); maturation (March 1983) and the peak (May 1983) of the 1982–83 El Niño and compares the anomalous conditions with observations made during November 1983, April 1984 and May 1977 when no thermal anomaly was present. Modified from ref. 8.

nitrate concentrations of $\geq 2 \mu\text{M}$ constitute favourable nutrient conditions where the uptake versus concentration relationship is saturated. Figures 3 and 4 show that, in the absence of El Niño, the equatorial and coastal waters of the eastern tropical Pacific are nutrient rich; that is, they contain saturating concentrations of nitrate in the surface layer. In contrast, during the 1982–83 El Niño the surface concentration of nitrate for six and a half months, from December 1982 to mid-June 1983, was very low, often below the detection limit of $0.02 \mu\text{M}$. Such concentrations are clearly nutrient-poor and will decrease the rate and quantity of new primary production by phytoplankton²³. The progression of changes in nitrate is paralleled by changes in phosphate and silicate; see ref. 18.

The November 1982 section at 5° S on the coast of Peru was sampled during onset of El Niño (Fig. 5), but the temperature anomaly had already reached 5°C by this time and thermocline and nutricline depression were evident. The 20°C isotherm, which is the middle of the thermocline, is depressed over 100 m close to the coast when compared with the section made in November 1983 (Fig. 5). Comparison of the surface temperature along the 5° S transect during November 1982 and November 1983 (Fig. 6) shows that the offshore to inshore temperature gradient was similar in magnitude even though the entire section is 5°C warmer during November 1982. The persistence of this onshore/offshore gradient indicates that upwelling continued close to the coast supplying the inshore waters band next to the coast with nutrients. During November 1982 the high phytoplankton biomass zone, defined by the 1 mg m^{-3} chlorophyll contour extended only ~30 km out from the coast; after recovery in November 1983 the high chlorophyll zone extended ~150 km from the coast (Fig. 5). There was an inshore productivity maximum present in November 1982, but productivity integrated across the 5° S transect from the coast to 85° W was much reduced (Fig. 6). The mean surface productivity along the transect during November 1982 was $51 \text{ mgC m}^{-3} \text{ day}^{-1}$, whereas

during November 1983 the average was $192 \text{ mgC m}^{-3} \text{ day}^{-1}$, or 3.8 times higher.

In March 1983, the thermocline was slightly shallower along the 5° S transect than it was during onset phase (Fig. 5) and coastal upwelling continued as shown by the upward tilt in the isotherms close to the coast and the continued presence of the offshore-to-inshore temperature gradient of ~3°C (Fig. 6). High chlorophyll concentrations in March 1983 were restricted to a 30-km wide zone next to the coast; after recovery in April 1984 the zone of high chlorophyll and high primary productivity extended 400 km from the coast. In March 1983, the mean surface primary productivity along the 5° S transect was reduced to $27 \text{ mgC m}^{-3} \text{ day}^{-1}$, or 9.2 times higher than the El Niño conditions.

May 1983 showed maximum expression of the biological and physical anomalies of the 1982–83 El Niño (Fig. 7). The surface temperature anomalies reached 10°C and the middle of the thermocline was depressed below 100 m (Fig. 5). During May 1983 the isotherms tilted downwards towards the coast, indicating onshore and poleward flow in the eastern boundary region rather than coastal upwelling. A comparable section in May 1977 showed the upward tilt typical of coastal upwelling with offshore and equatorwards flow. The chlorophyll distribution in May 1983 shows that the normally rich coastal ocean has changed to a relatively barren condition characteristic of a low latitude gyre, including even the narrow 30-km inshore zone of high chlorophyll water. However, by July 1983 (Fig. 6) a rapid recovery of normal upwelling resulted in the return of high phytoplankton productivity and biomass in the nearshore zone along the entire coast of Peru. During May 1983, the mean surface primary productivity along 5° S dropped to $10 \text{ mgC m}^{-3} \text{ day}^{-1}$, during July 1983 it was $219 \text{ mgC m}^{-3} \text{ day}^{-1}$, a value $21.3 \times$ higher than the El Niño condition. At the peak of the anomaly primary production in coastal waters was reduced to 5% of the quantity observed during non-El Niño conditions.

Effects on living resources

Large-scale and persistent changes in sea-surface temperature along the coast of Peru are well known because the warm anomalies have been associated with changes in the abundance of the Peruvian anchovy (Fig. 1). Early workers studying the effects of El Niño on fish and sea birds, particularly the Peruvian scientists who are most familiar with the phenomenon^{24,25}, believed that reductions were caused by reproductive failures resulting from disruption of the normal upwelling food chain; observations of nutrients and primary production during the 1982–83 El Niño confirm this connection. It seems logical that 4–20-fold reduction in productivity and biomass of phytoplankton would decrease the growth, survival and reproductive success of these organisms, but there are opposing views on the effect of El Niño on living resources. A decade ago it was suggested²⁶ that anchovies simply swam away to deeper, cooler water during the anomaly, and it was predicted in 1972 that the fish would return to the Peruvian coastal waters when the thermal anomaly ended; Fig. 1 shows that this prediction was not correct. After the warm anomalies of 1972 and 1976 anchovies did not return to their previous level of abundance and the catch became lower and lower.

These observations suggest that for some of the commercially important species of the upwelling ecosystem a major effect of El Niño is an absolute decrease in growth and reproduction; that is, the total productivity of the region was reduced, not simply rearranged in space along the coast. However, some dramatic large-scale redistributions of certain species of invertebrates and fish were observed during the 1982–83 event and it is important to consider both processes.

Redistribution is an accurate description of the response of hake (*Merluccius gayi*) to the 1982–83 El Niño. Fishermen report that the relatively large and motile bottom-dwelling hake moved down the continental slope, staying with the cool water to which it is adapted. Comparison of Fig. 5 with Fig. 7 shows that when the 16 °C isotherm was depressed to a depth of several hundred metres (November 1982), hake disappeared from the catch off northern Peru. The fishermen also indicate that in addition to going deeper hake moved or were transported southward²⁷. The stronger than average polewards flow of the Peru/Chile undercurrent¹⁶ may have been involved in this movement. It is not clear whether there was a temperature-mediated migration by hake southwards in response to the warm water or whether this and other species were passively transported by the strong southward flows in both surface²⁸ and subsurface currents¹⁶. Hake were the first fish to leave the northern Peru region but the last to return. In this context, it is clear that hake did not return as soon as the physical conditions in their habitat returned to normal. By November 1983 temperature conditions in the shelf and shelf-break environment had returned to normal (Fig. 5), but despite continued fishing effort, significant catches of hake did not resume until January 1984 (Fig. 7). The hake yield in late 1984 at northern Peru was high²⁷, suggesting that the fish returned in both large numbers and relatively good physiological condition.

The increase in shrimp catches at Paita (Fig. 7) is an accurate index of changes in abundance that occurred throughout the region from Colombia to Peru²⁹. The catch of the three major species of shrimp (*Xiphopenaeus riueti*, *Penaeus occidentalis* and *Trachypenaeus hyesi*) did not change on the Pacific coast of Colombia from January to September 1982 but decreased in the last three months of that year. The shrimp harvest along the coast of Ecuador increased greatly during the first half of 1983 (ref. 29) and the catch in northern Peru, which is usually very low³⁰, began to increase as the temperature increased from November 1982 to February 1983 (Fig. 7). The southwards progression of shrimp abundance probably resulted from their partially planktonic character. That is, shrimp may have been carried southward by the nearshore current that is historically associated with El Niño³¹. It is unlikely that the southwards

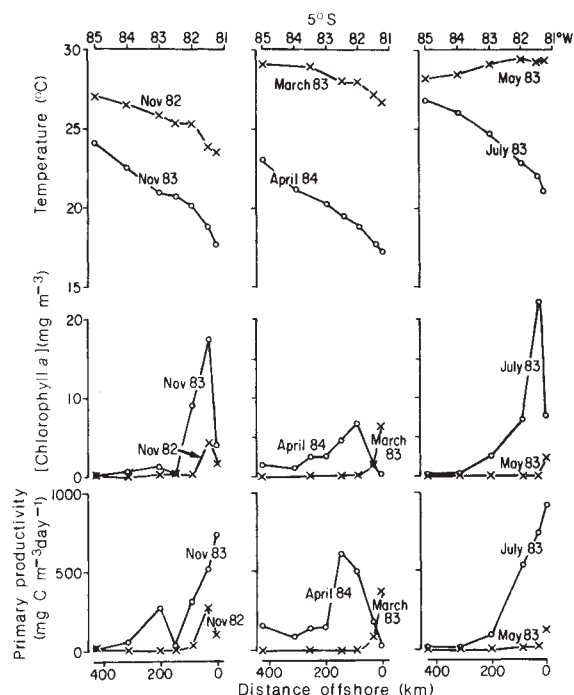


Fig. 6 Surface temperature, chlorophyll and primary production along 5° S from 85° W to the coast of Peru showing the comparison between El Niño (November 1982, March 1983 and May 1983) conditions and conditions during November 1983, April 1984 and July 1983 when no thermal anomaly was present.

redistribution was the result of an organized behavioural migration because penaeid shrimp are relatively weak swimmers.

The scallop *Argopecten purpuratus* underwent a rapid increase in abundance during the anomaly; in 1983 the harvest of scallops increased 40-fold to ~20,000 tonnes (ref. 32). By January 1984, the population increase was over and the abundance had returned to its previous level³². Scallops are common permanent residents in the inshore waters of Peru, so the increase did not necessarily result from migration or transport. Adult scallops are sessile and their planktonic larvae are not thought to survive long distance transport, so it is hypothesized that the extraordinary population increase resulted from *in situ* increases in growth and reproduction. Positive effects from the elevated temperatures that may have contributed to the population increase are: (1) faster maturation of the adults; (2) faster larval development causing reduced dispersion of the planktonic stages; (3) reduction of predators; and (4) faster maturation of the juvenile scallops after they settle out of the plankton. Because a generation of scallops requires from 4 to 8 months in most favourable conditions this increase was apparently the result of one or two generations of scallops that enjoyed extremely high reproductive success and then found excellent growing conditions. Although there was a general decrease in the abundance of phytoplankton food in the region (Fig. 6), the quantity of phytoplankton available during the anomaly must have been adequate for rapid growth and successful reproduction. An inshore maximum of chlorophyll and primary production persisted along the Peru coast even at the peak of the anomaly in May 1983 (see Fig. 6). Another species of the genus *Argopecten*³³ showed a saturation in the growth rate at 28 °C requiring a chlorophyll concentration of 2.4 mg m⁻³; the Peruvian scallop beds contained chlorophyll concentrations in this range in February 1983³⁴. Although the phytoplankton biomass was reduced from the climatological mean, it remained high enough in the nearshore waters to support a dramatic, but very short-lived, population increase in this species.

Jack mackerel (*Trachurus symmetricus*) disappeared from the catch in the Paita region in December 1982 (Fig. 7) between the periods that hake and sardine disappeared from the catch.

This timing is closely connected to the progression of physical changes in the habitats of these three species. Jack mackerel normally occupy the offshore half of the coastal upwelling ecosystem where their euphausiid food is most abundant⁸. The warm ($>26^{\circ}\text{C}$) water present in the outer 300 km during November 1982 (Fig. 5) concentrated jack mackerel in the inner 100 km even during onset. Oceanic predators in the form of bonito (*Sarda chiliensis*), dorado (*Coryphaena hippurus*) and yellowfin tuna (*Thunnus albacores*) became much more abundant during onset of the 1982–83 El Niño³⁰. Increased predation on the concentrated jack mackerel by both man and natural predators must have contributed to the reduction in abundance December 1982 in northern Peru, but this species also apparently widened its latitudinal distribution to both the north and the south. The catch of jack mackerel off Chile decreased from 1.5 million tonnes in 1982 to 0.8 million tonnes in 1983. In late 1983, jack mackerel returned in relatively large number to the northern Peru region (Fig. 7) and in 1984 catch off Peru and Chile returned to the pre-Niño level³⁵. The fish were reported to have relatively normal length-weight relationships indicating that jack mackerel dispersed to areas of favourable food conditions. Clearly, this species has a remarkable ability to survive even a very severe El Niño.

In 1983, the sardine (*Sardinops sagax*) decreased almost to zero in the Ecuadorian catch, decreased slightly in the Peruvian catch (Fig. 1) and increased in the Chilean catch, indicating a southwards shift of the population^{35,36}. Despite this large-scale southward shift, the maintenance of relatively high levels of sardine abundance off northern Peru throughout March 1983 (Fig. 7) was consistent with environmental conditions revealed by cross-shelf profiles (Fig. 5). During the November 1982 to March 1983 period sardines remained in the narrowing band along the coast where chlorophyll concentrations were $>1\text{ mg m}^{-3}$. Chlorophyll is an index of phytoplankton abundance, but because phytoplankton are the major food of zooplankton, chlorophyll concentration can also delimit habitat where both zooplankton and phytoplankton are abundant. In addition, stomach content analyses carried out in Paita in 1983 showed that 47% of the material in sardine stomachs was phytoplankton and 40% was zooplankton. These observations suggest that chlorophyll concentration is an approximate index of habitats that are favourable to sardines.

As adult sardines concentrated in the nearshore region they were easily caught, but reduction of their oil content to $<1\%$ by weight and reduction in individual body weight by 20% (ref. 36) indicated that their nutritional condition was decreasing. By April 1983 (Fig. 7), sardines disappeared from the Paita region and the May 1983 profiles show that water with $>1\text{ mg m}^{-3}$ chlorophyll was virtually eliminated from the region and primary production was reduced 20-fold (Fig. 6). The reduction in sardine catch at Paita in May 1983 was indicative of the reduction that occurred along the entire coast of Peru³⁶ during that period and Fig. 1 shows that the catch for all of 1983 was also reduced.

The most dramatic biological effect of the 1982–83 El Niño was on the Peruvian anchovy (*Engraulis ringens*), once the basis of the world's largest fishery^{24,25}. Figure 1 shows the covariation of thermal anomalies and anchovy catch; we suggest that this correlation reflects a causal relationship that depends on changes in the large-scale thermal and nutrient dynamics that take place during El Niño in the Pacific. Depressing the nutricline (Fig. 3) makes nutrient conditions less favourable (Fig. 4) for phytoplankton primary production. The 20-fold decrease in phytoplankton production that occurred during the peak of the anomaly in 1983 necessarily decreased growth, survival and reproductive success of the anchovy. Growth and survival were affected because adult anchovy feed, in part, directly on phytoplankton³⁸. Because larval survival is a function of the abundance and type of phytoplankton food³⁷, reproductive success was especially reduced in the 1982–83 event³⁶. Taken together,

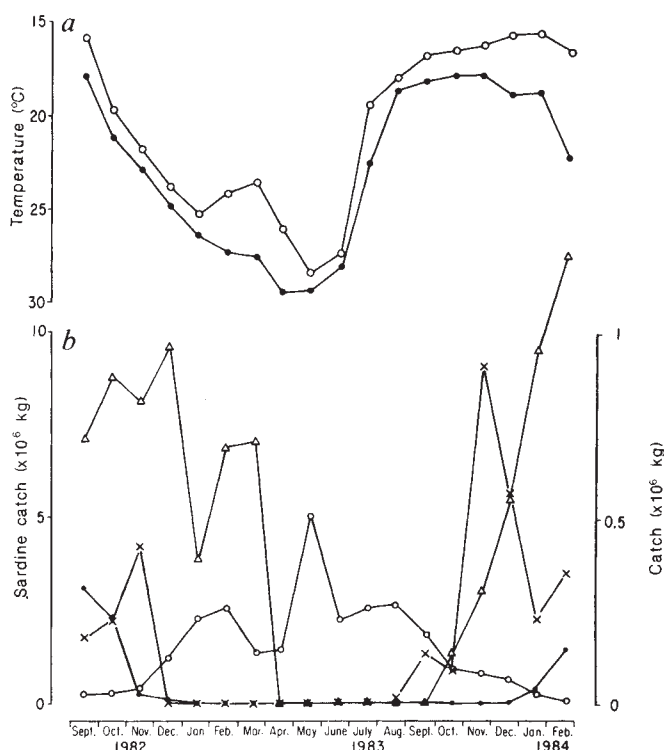


Fig. 7 a, Monthly mean surface (●) and 60-m (○) temperatures at an ocean station off Paita, Peru (5°S). b, Monthly landings of sardine (Δ), jack mackerel ($\times$), hake (●) and shrimp (○) at Paita. Updated from ref. 8.

these larval and adult processes reduced the anchovy stock to record low levels during the 1982–83 El Niño. The catch in 1983 was 118,000 tonnes $<1\%$ of the catch of a decade ago (Fig. 1).

Of the important fish stocks in this region (hake, mackerel, jack mackerel, sardine and anchovy), anchovy experienced the greatest mortality during the 1982–83 El Niño. This vulnerability can be explained in general terms if the anchovy is the species best adapted to the high biological productivity of coastal upwelling. Being best adapted to high productivity, it is the species most affected by a reduction in the productivity. The vulnerability can also be explained in terms of a specific behavioural mechanism of anchovy. Anchovy seek out and remain in relatively cool water of $16\text{--}18^{\circ}\text{C}$ (ref. 38). During normal conditions this behaviour leads anchovy to the portion of the coastal upwelling ecosystem where phytoplankton are most abundant (see the location of the chlorophyll maximum in Fig. 6). During onset of El Niño this behaviour causes anchovy to concentrate in the persistent inshore upwelling; the distribution of these relatively cooler centres along the coast as late as February and March 1983 was shown by Smith and Huyer¹⁶. Concentration in these upwelling centres is adaptive behaviour if the El Niño is weak or does not fully develop, as was the case in 1975 (ref. 39); but during a very strong El Niño this behaviour makes the persistent upwelling centres traps for anchovy. As the 1982–83 event progressed, anchovy that concentrated³⁶ in the upwelling centres were surrounded by warmer water on three sides and land on the fourth side. At this stage in the development of the anomaly, temperature-mediated behaviour was unable to lead the fish out of the upwelling pockets to more favourable habitat. When the cool pockets were eliminated by alongshore or onshore flow large numbers of anchovies perished⁸.

An indication of the severity of nutritional stress on the Peruvian anchovy was seen in the effects of the 1982–83 El Niño on the northern anchovy (*Engraulis mordax*) off California where expression of the anomaly was similar to that off South America, but considerably weaker. Off California, there were thermocline anomalies⁴⁰, reversal of flow to a poleward direc-

tion⁴¹ and reduced phytoplankton and primary production^{42,43}. Analysis of the response of northern anchovy⁴⁴ to the anomaly showed that: (1) body weight of spawning females was very low in 1983–84; (2) spawning occurred at smaller sizes than ever before observed; (3) growth of juveniles was 36% less than previous years; and (4) the spawning biomass was the lowest since 1962. Because the reduction in plankton biomass and productivity was proportionally much greater in the habitat of the Peruvian anchovy, the increased reproductive failure and adult mortality off America is not surprising.

Abrupt changes in the dominance of pelagic fish stocks after decades of apparent stability have long been unexplained⁴⁶. Steele and Henderson⁴⁷ explained such changes by using a population model with multiple equilibrium states and subjecting the model to environmental variability that increased in variance per unit frequency as periodicity increased from years to decades. The extreme variance of 1982–83 El Niño is a classic example of the reality of the environmental assumption used by Steele and Henderson⁴⁷, but El Niño and longer-term climate anomalies have always been part of anchovy's environment. A dominance shift as a result of the 1982–83 El Niño would provide evidence that fish stocks are more vulnerable to environmental variability when they are heavily exploited.

Since the end of the 1982–83 event, environmental conditions

in the coastal upwelling ecosystem and phytoplankton productivity have returned to normal (Fig. 6). Indeed, the levels of primary production observed in July 1983, November 1983 and April 1984 were among the highest observed in 19 yr along the Peru coast⁹. The resilience of the species that are major living resources in the region affected by the 1982–83 El Niño is made more interesting by the rapid return of very high levels of phytoplankton productivity. Visual observations made on a March 1985 cruise and reports from fishermen indicate that relatively large adult anchovies are once again abundant in certain locations along the coast. These reports suggest anchovy possess considerable resilience to El Niño even when heavily exploited. In this context, anchovy appears to be the species best adapted to the alternating El Niño/upwelling cycle; the enormous reproductive capacity of this species^{24,25,38} enables it to recolonize the upwelling ecosystem faster than other pelagic fish after a very severe anomaly. Although the evidence is incomplete, there are indications that the living resources along the west coast of South America are in the process of rapidly recovering from the 1982–83 El Niño.

This research was supported in part by NSF grant OCE-8110702. Information on monthly fish catch at Paita was provided by Instituto del Mar del Peru. We thank colleagues in Peru, Ecuador and Chile for providing data.

1. Philander, S. G. H. *Nature* **302**, 295–301 (1983).
2. Gill, A. E. & Rasmusson, E. M. *Nature* **306**, 229–234 (1983).
3. Rasmusson, E. M. *Am. Scient.* **73**, 168–177 (1985).
4. Wormuth, J. H. & Berkowitz, S. P. *Eos* **65**, 922 (1984).
5. Shulenberg, E. & Loeb, V. J. *Eos* **65**, 922 (1984).
6. Heywood, R. B., Everson I. & Priddle, J. *Deep-Sea Res.* **32**, 369–378 (1985).
7. Ryther, J. H. *Science* **166**, 72–76 (1969).
8. Barber, R. T. & Chavez, F. P. *Science* **222**, 1203–1210 (1983).
9. Barber, R. T. & Smith, R. L. in *Analysis of Marine Ecosystems* (ed. Longhurst, A. R.) 31–68 (Academic, New York, 1981).
10. Barnett, T. P. *J. phys. Oceanogr.* **11**, 1043–1058 (1981).
11. Wyrtki, K. J. *J. phys. Oceanogr.* **5**, 572–584 (1975); *J. geophys. Res.* **89**, 10419–10424 (1984).
12. Mangum, L. J. & Hayes, S. P. *J. geophys. Res.* **89**, 10441–10449 (1984).
13. Cane, M. A. *Science* **222**, 1189–1195 (1983).
14. Rasmusson, E. M. & Wallace, J. M. *Science* **222**, 1195–1202 (1983).
15. Enfield, D. B. *J. geophys. Res.* **86**, 2005–2016 (1981).
16. Smith, R. L. *Science* **221**, 1397–1399 (1983).
17. Smith, R. L. & Huyer, A. *Trop. Ocean-Atmos. Newslett.* **21**, 28–29 (1983).
18. Barber, R. T., Kogelschatz, J. E., Chavez, F. P. & Thayer, V. G. in *El Niño Atlas 1982–1983* (eds Leetma, A. & Witte, J.) 149–155 (US Government Printing Office, Atlanta, 1984).
19. Lukas, R., Hayes, S. P. & Wyrtki, K. J. *J. geophys. Res.* **89**, 10425–10430 (1984).
20. Chavez, F. P., Barber, R. T. & Soldi S., H. *Nature* **309**, 47–49 (1984).
21. Dugdale, R. C., Jones, B. H., Jr, MacIsaac, J. J. & Goering, J. J. *Can. Bull. Fish. Aquat. Sci.* **210**, 234–250 (1981).
22. Eppley, R. W. *Can. Bull. Fish. Aquat. Sci.* **210**, 251–263 (1981).
23. Dugdale, R. C. *Limnol. Oceanogr.* **12**, 685–695 (1967).
24. Guillen, O., Calienes, R. & Izaguirre de Rondon, R. *Bol. Inst. Mar Peru* **2**, 49–76 (1969).
25. Valdiva G., J. E. *Rapp. P. V. Reun. Cons. int. Explor. Mer* **173**, 196–202 (1978).
26. Murphy, G. T. *Geoforum* **11**, 63–71 (1972).
27. Arntz, W. E. *Oceanus* **27**, 36–39 (1984).
28. Castillo, R. *Estudio Regl Fenommon El Niño Bol.* **12**, 3–5 (1985).
29. Halpern, D., Leetma, A., Hansen, D. & Philander, S. G. *Science* **221**, 1173–1174 (1983).
30. Herdson, D. *Trop. Ocean-Atmos. Newslett.* **28**, 14–16 (1984).
31. Velez, J., Zeballos, J. & Mendez, M. *Trop. Ocean-Atmos. Newslett.* **28**, 10–12 (1984).
32. Eguiguren, V. *Bol. Soc. Geogr. Lima* **4**, 241–258 (1894).
33. Wolff, M. *Trop. Ocean-Atmos. Newslett.* **28**, 8–9 (1984).
34. Kirby-Smith, W. W. & Barber, R. T. *Aquaculture* **3**, 135–145 (1974).
35. Barber, R. T. & Kogelschatz, J. E. *Data Rep. of WECOMA Cruises WELOC 83-2 and 83-3* (Duke University, Beaufort, USA).
36. Avaria, S. *Estudio Regl Fenommon El Niño Bol.* **13**, 17–25 (1985).
37. Santander, H. & Zuzunaga, J. *Trop. Ocean-Atmos. Newslett.* **28**, 9–10 (1984).
38. Lasker, R. *Fish. Bull.* **73**, 453–462 (1975).
39. Jordan, R. *Invest. Pesq.* **35**, 113–126 (1971).
40. Cowles, T. J., Barber, R. T. & Guillen, O. *Science* **195**, 285–287 (1977).
41. Simpson, J. J. *J. Geophys. Res. Lett.* **10**, 937–940 (1983).
42. Lynn, R. J. *J. geophys. Res. Lett.* **10**, 1093–1095 (1983).
43. Fieldler, P. C. *Science* **224**, 1251–1254 (1984).
44. McGowan, J. A. *Trop. Ocean-Atmos. Newslett.* **21**, 23 (1983).
45. Fiedler, P. C. *CalCOFI Rep.* **25**, 53–58 (1984).
46. Fiedler, P. C. & Hewitt, R. P. *Eos* **65**, 910 (1984).
47. Steele, J. H. *Nature* **313**, 355–358 (1985).
48. Steele, J. H. & Henderson, E. W. *Science* **224**, 985–987 (1984).

ARTICLES

Global resonances in the evolution of solar magnetic fields

J. O. Stenflo & M. Vogel

Institute of Astronomy, ETH-Zentrum, CH-8092 Zurich, Switzerland

Decomposition of the pattern of solar magnetic fields in spherical harmonics for a data set of 25 years and power spectrum analysis of the harmonic coefficients reveals a strikingly resonant modal structure. The resonance frequencies contain information on the structure of the magnetic fields in the Sun's interior.

THE remarkably rich spectrum of acoustic global oscillations of the Sun with periods around 5 min (refs 1, 2) demonstrates how an apparently chaotic body in a highly turbulent state can exhibit extremely well-defined resonant behaviour. Exploration of the resonance frequencies allows a determination of the internal structure of the Sun².

Such resonances naturally result in any closed system, when perturbations around an equilibrium can be described by a wave

equation. A spherical system like a star or an atom forms closed cavities in which the waves can interfere constructively or destructively. This leads in the case of the de Broglie waves in the atoms to the resonant structure exhibited by atomic spectra. Analysis of such spectra allows the internal structure of matter to be determined. Analogously, the acoustic waves in the Sun give rise to the p-mode oscillation spectrum. Because the orthogonal functions used to describe the eigenmodes in

spherical geometry are the spherical harmonics, characterized by the n , l and m quantum numbers, the classification of both stellar and atomic modes is in terms of these numbers.

In the case of the Sun, the wave equation describes how the perturbation of the equilibrium state tends to be restored under the action of a force. When the restoring force is the pressure force, the spectrum of p-mode oscillations results. When the restoration is done by the gravity force, another spectrum is obtained. Attempts have been made to observe this g-mode spectrum³⁻⁵, which has oscillation periods of the order of 1 h or longer, but the g modes have not yet been definitely identified empirically.

There is a third force on the Sun that has not yet been invoked in this connection: the electromagnetic force. If, for instance, waves of hydromagnetic nature (such as dynamo waves) could interfere constructively and destructively globally in the Sun, a third type of resonant spectrum could result. The expected resonance periods depend on the strength and depth distribution of the magnetic fields inside the Sun, but, as we shall see, crude estimates indicate periods much longer than those of the p or g modes, typically in the range of years.

Sunspot magnetic polarities vary with a magnetic cycle of 22 yr, a cycle that the polarities of the polar magnetic fields on the Sun also seem to follow, although the north and south heliographic poles do not reverse polarity in synchrony⁶. The whole solar surface is, however, covered by a complex magnetic field pattern, which is highly structured from the largest to the smallest (sub-arc second) scales, and which evolves all the time. Although this evolution is very complicated, it seems to proceed in an orderly manner, as though globally regulated.

Attempts have been made to explain the 22-yr magnetic cycle and the migration of the sunspot latitude zones in terms of dynamo theory⁷⁻⁹, but this approach has recently encountered difficulties. No attempt has been made to see if the evolution of the global magnetic-field pattern can be explained in terms of constructive and destructive interference of waves.

We demonstrate here that the magnetic-field pattern, in fact, has a modal structure characterized by sharp global resonances. Furthermore, the resonances for the modes of odd and even parity are decoupled from each other, indicating an underlying selection rule. This new 'emission-line' spectrum of solar magnetic fields does not seem to be explicable within the framework of current concepts such as dynamo theory, but should contain potentially powerful diagnostic information on the interior magnetic structure of the Sun and on the origins of solar activity.

Data set and reduction procedure

To explore whether there are global resonances in the Sun's magnetic field, we have to decompose the field pattern in its spherical harmonic coefficients and then perform a power spectrum analysis of each coefficient, showing the results in frequency (ν)- l space (as for the p-mode oscillations). As the expected periods are very long (years), we need synoptic observations of the magnetic field over all heliographic latitudes and longitudes, over a time span of many years, if possible longer than the 22-yr magnetic cycle. Thanks to the synoptic magnetogram programme started at the Mount Wilson Observatory in 1959 and at the Kitt Peak National Observatory (KPNO) in 1975, such a data set is available. The combined data set that we have used thus covers 25 yr.

Each synoptic map of the magnetic field, giving the average field strength as a function of heliographic latitude and longitude, has been pieced together from daily magnetograms obtained over one rotation period (about 27 days) of the Sun. Normally, the field observed around the central meridian in each magnetogram is used to provide the data for the various longitude bands. When some days of observations are missing, magnetogram data away from the central meridian are used to fill in the various longitudes. The longitudes are assigned according to the system introduced by R. Carrington, based on a

rotation period of 27.2753 days. This period may seem somewhat arbitrary, since the Sun rotates differentially, but the precise choice of longitude system does not affect the type of analysis that we performed. By convention, the Carrington periods are numbered with the first period starting 9 November 1853. Our full data set covers rotations 1,417-1,751, corresponding to August 1959-July 1984.

The data from Mt Wilson cover rotations 1,417-1,648 (August 1959-December 1976), the Kitt Peak data rotations 1,622-1,751 (January 1975-July 1984). The major data gaps are for Mt Wilson rotations 1,462-1,468 (January-June 1963), for Kitt Peak rotations 1,640-1,644 (April-July 1976) and 1,647-1,648. As the latter data gaps are covered by the Mt Wilson data, we use the Kitt Peak data only from rotation 1,649 when both data sets are connected together as a continuous, 25-yr time series. When, on the other hand, the Kitt Peak data are used by themselves (as when we want to exploit their higher spatial resolution), the full series starting with rotation 1,622 is used. Each synoptic map is a matrix of values of the line-of-sight component of the field strength averaged over intervals in latitude and longitude. The divisions are equidistant in longitude and in the sine of the latitude. The Mt Wilson data that we have obtained were transformed into the synoptic format largely by Altschuler *et al.*¹⁰ at the High Altitude Observatory, and have 30 zones in sine(latitude) and 36 sectors (10° each) in longitude. The Kitt Peak data compiled by J. W. Harvey has the higher-resolution format of 180 zones and 360 sectors.

The raw synoptic maps are uncorrected for actual inclination of the field vector, Zeeman saturation and temperature weakening of the spectral line used for the observations, and magnetogram cutoff for strong sunspot magnetic fluxes. The correction for these effects of the Mt Wilson data has been detailed elsewhere¹¹, so we will only summarize the procedure here.

To correct for the inclination of the magnetic field to the line-of-sight, we have to assume what the true field direction is. A good approximation should be to assume that the magnetic field in the photospheric layers, where it is observed, is, on average, along the radial direction; the reason for this is that the large-scale magnetic flux is made up of very small-scale, discrete flux tubes, and that buoyancy forces acting on each of them tend to make them as vertical as possible.

In their computations of the extrapolated coronal magnetic fields, Altschuler and Newkirk¹² postulated that the direction of the photospheric magnetic fields is determined by the solution of Laplace's equation for the assumed vacuum field in the solar corona, with the observed photospheric line-of-sight component as a lower boundary condition. This assumption, however, cannot be valid, since the force-free approximation is not applicable in the photosphere, where the field is measured. Instead, the photospheric field consists of highly intermittent flux tubes, which are not affected by the global topology of the coronal field because very local pressure and gravity forces exceed magnetic tension forces. Higher in the atmosphere the field must adjust to the force-free state, but that is not where the field is measured.

Accordingly, assuming that the photospheric field is radial on average, we have transformed all the synoptic charts into maps of the radial magnetic flux. As the Mt Wilson synoptic maps become saturated at about 100 G, flux densities in sunspots will be greatly underestimated unless this effect is corrected for. Therefore, for each synoptic map, a set of sunspot corrections has been applied¹¹. The Kitt Peak measurements do not contain this type of saturation, so the sunspot fluxes measured are less distorted. Accordingly, no sunspot correction has been applied to the Kitt Peak data.

The Mt Wilson magnetograph recordings have been made with the spectral line Fe I $\lambda 5,250.22$ Å, which has considerable Zeeman saturation (nonlinearity in the polarization signal when the Zeeman splitting is large) and temperature sensitivity (weakening of the line in magnetic regions). These effects and

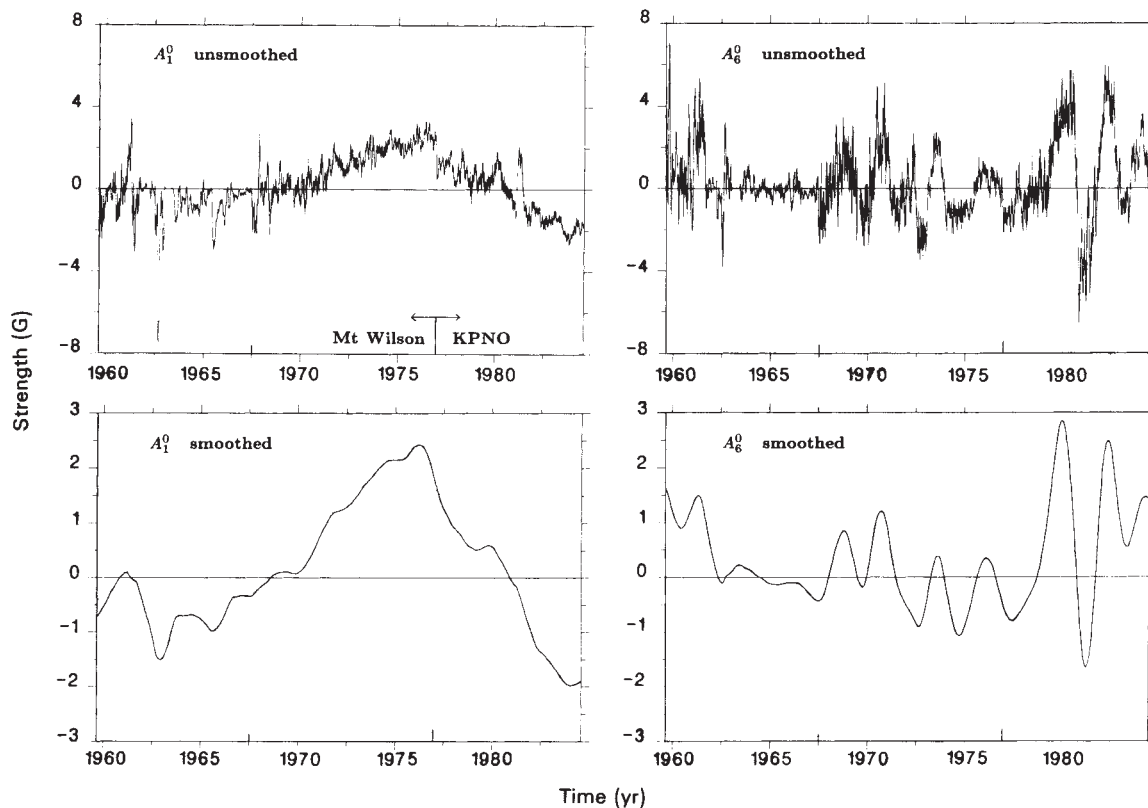


Fig. 1 Time series of spherical harmonic coefficients A_1^0 and A_6^0 . The upper diagrams show the unsmoothed data, the lower diagrams the curves when smoothed with a triangular window of total half-width 1 yr.

their centre-to-limb variations have been calibrated¹³, so we have been able to correct all the Mt Wilson synoptic charts for them. The Kitt Peak observations, on the other hand, are based on the spectral line Fe I $\lambda 8,688.64\text{\AA}$, which has much less temperature sensitivity and Zeeman saturation. As in addition, no calibrated correction factors for this line are available, the Kitt Peak data have not been corrected for these spectral effects, which should in any case be small.

Expanding the radial field B_r in spherical harmonics, and redefining the coefficients to make them real and to absorb in them time-invariant factors, we obtain

$$B_r(\vartheta, \varphi, t) = \sum_{l=0}^{\infty} \sum_{m=0}^l [A_l^m(t) \cos m\varphi + B_l^m(t) \sin m\varphi] P_l^m(\cos \vartheta) \quad (1)$$

where P_l^m are the associated Legendre polynomials, ϑ is the co-latitude, φ the longitude and t the time. The coefficients can be obtained from the observed B_r through

$$\begin{aligned} A_l^0(t) &= \frac{2l+1}{4\pi} \int B_r(\vartheta, \varphi, t) P_l(\cos \vartheta) d\Omega, \\ \left\{ \begin{matrix} A_l^m(t) \\ B_l^m(t) \end{matrix} \right\} &= \frac{2l+1}{2\pi} \frac{(l-m)!}{(l+m)!} \\ &\times \int B_r(\vartheta, \varphi, t) P_l^m(\cos \vartheta) \begin{Bmatrix} \cos m\varphi \\ \sin m\varphi \end{Bmatrix} d\Omega \end{aligned} \quad (2)$$

where the integrations are to be made over all solid angles. In practice, the integrals are, of course, replaced by sums over the discrete matrix representing the synoptic chart.

In the reduction, the small unphysical magnetic monopole contribution is calculated by

$$A_0^0 = \frac{1}{4\pi} \int B_r d\Omega \quad (3)$$

and subtracted from the original B_r values. These corrected B_r values are used in equation (2) to compute the A_l^m and B_l^m coefficients.

Legendre polynomial coefficients of the Sun's magnetic field have been derived previously by others^{10,12,14-16} by fitting a solution of the Laplace equation for the coronal field to the observed photospheric line-of-sight field. This procedure is not used here, for three reasons: (1) It is physically incorrect (the Laplace equation has no relevance to the photosphere, as pointed out above). (2) All the coefficients become coupled to each other, so the values of the low-degree coefficients depend on the cutoff of the expansion, with poor convergence properties. (3) The computational speed of our method is of the order of 1,000 times faster.

To obtain a practically continuous time series for each of the coefficients, the consecutive synoptic charts have been shifted relative to each other not by one solar rotation, but by 10° in longitude. The time step in the series is thus $27.2753/36 \approx 0.76$ days. We recall that the observational sampling was made at daily intervals, but that a synoptic chart is composed of observations collected over 27 days. The exploration of evolutionary effects is thus limited to longer periods.

A 25-yr time series thus consists of about 12,000 data points. The next step is to compute power spectra $PA_l^m(\nu)$ and $PB_l^m(\nu)$ of $A_l^m(t)$ and $B_l^m(t)$, where ν is the frequency. To obtain a smoother representation of the power spectra when examining the low-frequency domain, the time series were extended with zeroes before computing the fast Fourier transforms, thereby reducing the spacing between the discrete ν values in the power spectra.

The spatial resolution determined by the discretization in latitude and longitude of the synoptic charts allows the coefficients to be calculated up to l -values of 14 in the case of the Mt Wilson data and 89 in the case of the Kitt Peak data. All our computations have been made with a VAX 11/780 and an image processing system De Anza IP-8500 at the Institute of Astronomy in Zurich.

Resonant structure of zonal modes

We will limit ourselves here to a discussion of the results for the zonal modes, those with spherical harmonic order $m=0$, which are symmetrical around the rotational axis. The properties of the other modes will be considered elsewhere.

A hint at the resonant structure of the magnetic-field pattern is obtained from the time series of the harmonic coefficients. Figure 1 illustrates the time variation of coefficients A_1^0 (the north-south magnetic dipole) and A_6^0 . The portions of the time axis covered by the Mt Wilson and Kitt Peak data are indicated. In June 1967, a significant improvement of the Mt Wilson magnetograph system was introduced^{10,17,18} (as marked by an upward-directed tick mark in Fig. 1). Before then, the quality of the data was poorer. The smoothing in the lower half of Fig. 1 has been done using a triangular window with a total half-width of 1.0 yr, to bring out the evolutionary characteristics.

Hoeksema¹⁶ calculated the A_1^0 coefficient for the time span 1976–83 using a different method, but with results in qualitative agreement with ours.

The dominating feature of the north-south dipole is its expected variation with a 22-yr period, with sign reversals near the times of maximum solar activity. The behaviour of the coefficient A_6^0 is, however, qualitatively entirely different. The 22- or 11-yr cycles cannot be seen, but instead there is a very pronounced cyclic variation with a period of 2 yr. Inspection of the time series of other coefficients with even l values shows a similar cyclic behaviour, but the periods depend on l , whereas coefficients with odd l only show the 22-yr cycle (when $l \leq 14$).

Only a power spectrum analysis, however, can properly bring out the remarkable resonant structure of the magnetic-field pattern in an objective way. Figure 2 plots the power $PA_l^0(\nu)$ against frequency ν and harmonic degree l , for $l \leq 14$ and $\nu \leq 30$ nHz. Since the modal structure is found to be entirely different for odd and even l , we have plotted the results for the odd and even modes separately. The open box to the right of the diagrams gives the resolution in ν - l space. The frequency resolution is given by $1/T$, where T is the length of the time series, in our case 25 yr.

The modes with odd l are antisymmetrical, those with even l symmetrical, with respect to reflections in the plane of the equator. They therefore represent modes of odd and even parity in the magnetic-field pattern.

Figure 2 represents the combined 25-yr data set Mt Wilson-Kitt Peak. To form a homogeneous set, the higher spatial resolution in the Kitt Peak data has been degraded by averaging over the latitude bins defining the resolution elements in the Mt Wilson synoptic maps. We have further normalized the power to its maximum value separately for each value of l . This normalization, which also applies to Figs 3 and 4, shows more clearly the frequency structure in the grey-scale diagrams.

The most striking feature of Fig. 2 is that the power is concentrated at discrete frequencies, with the modes of odd and even parities forming two decoupled systems. The power for the odd modes is concentrated around a sharp resonance with a period of 22 yr, independent of the value of l . This corresponds to the well-known 22-yr cycle of sunspot polarities (twice the cycle of the sunspot number). It is, however, remarkable that no overtones (such as the 11-yr period) are present.

With the exception of $l=10$ and possibly $l=14$, the 22-yr cycle is absent in the even-parity modes. Instead, there is another well-defined resonance for each value of l . The resonance frequency increases in a systematic way with l , as we have emphasized by drawing a smooth curve through the maxima in Fig. 2b. Thus the resonant period is about 14 yr for $l=2$ and 1.4 yr for $l=14$. This ν - l relation for the even-parity modes is qualitatively similar to the ν - l diagram for the p-mode oscillations, suggesting that the underlying mechanism is a system of standing waves in the Sun's interior, although other interpretations might be possible. Whatever the cause, this new 'emission-

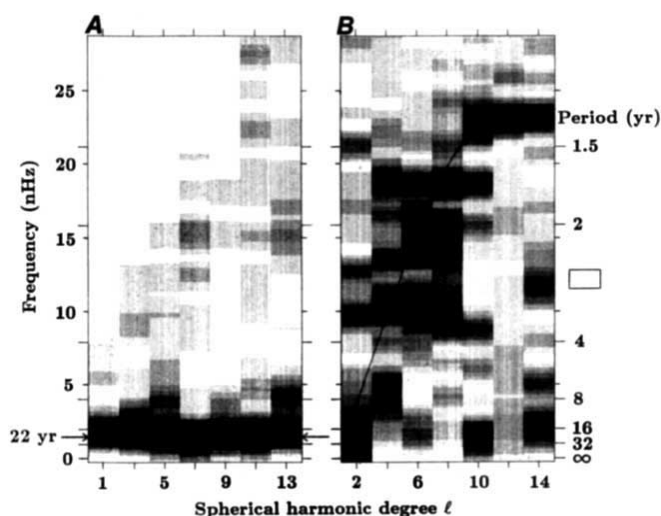


Fig. 2 Power spectra of the harmonic coefficients A_l^0 , as a function of frequency in nHz and spherical harmonic degree l , for $l \leq 14$. A, Odd values of l ; B, even values. A smooth curve has been drawn in the diagram for even l to indicate the l dependence of the resonance frequencies. The resolution is shown by the open box to the right.

line' spectrum is likely to contain information on the magnetic field distribution inside the Sun.

The resolution box to the right in Fig. 2 indicates that the frequency widths of the resonances are still unresolved. The coherent time of the standing waves thus seems to be longer than 25 yr.

Let us next consider the reliability of these results and how they are influenced by various spurious effects. To do this, we use the data in Fig. 3 to address the following questions. (1) Does the limited spatial resolution of the Mt Wilson data lead to aliases from higher spatial frequencies, which are under-sampled? (2) Do the Mt Wilson and Kitt Peak data, when treated separately, give the same results, although the time periods and instrumentation are different? (3) How stable are the results with respect to the length of the time series?

Figure 3 compares the ν - l diagrams of three data sets for $l \leq 24$: a, the Kitt Peak data, using their full spatial resolution of 180 latitude zones; b, the Kitt Peak data, spatially degraded to the Mt Wilson format by averaging over the 30 latitude bins (as was done when they were combined with the Mt Wilson data in Fig. 2); c, the Mt Wilson data. The resolution for the respective data sets is given in Fig. 3A for the Kitt Peak data (open box in Fig. 3A b) and the Mt Wilson data (Fig. 3A c).

Let us first consider the problem of aliasing. For a given harmonic component, there are $l-m+1$ latitude zones, that is, $l+1$ zones in our case (since $m=0$). To avoid aliasing, one needs to sample twice per zone. Thus, with 30 latitude bins as in the Mt Wilson format, one would expect to be able to determine modes up to $l=14$ without aliasing, if the zones were equidistant in the sine of the latitude (as our bins are). Since, however, the zones are not spaced this way, aliasing may also result for $l \leq 14$, depending on the importance of the high-latitude magnetic fields, for which undersampling due to the difference between the bin and zone divisions becomes significant. On the other hand, the averaging of the magnetic field over each bin in sine latitude acts as a spatial filter, suppressing the high-frequency components, which could be potentially responsible for aliasing.

Figure 3 shows that the aliasing and averaging effects are insignificant in the region $l \leq 14$, where the degraded and undegraded Kitt Peak data give practically identical results. This gives us confidence that Fig. 2 (which contains the results for $l \leq 14$) is free from limited spatial resolution effects. For $l > 14$,

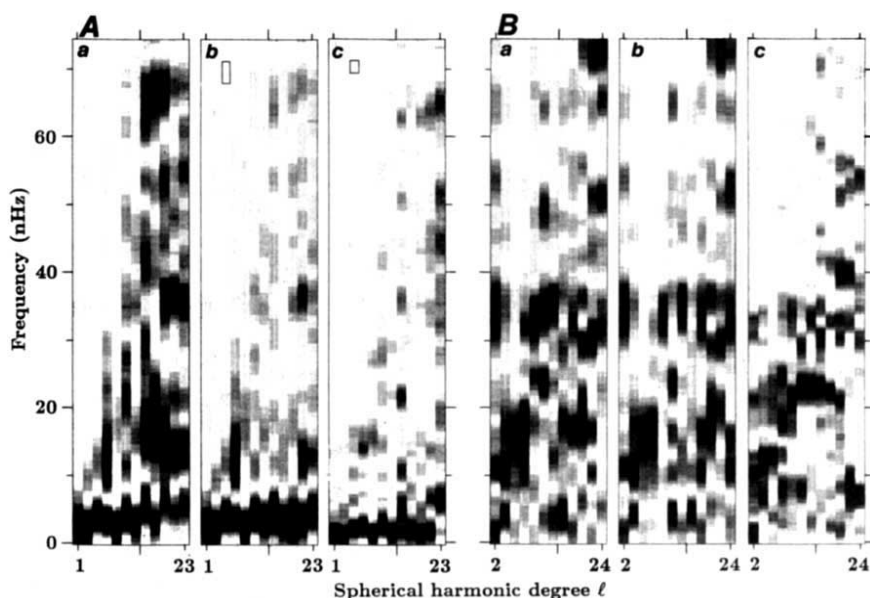


Fig. 3 Comparison of the modal structures for $l \leq 24$ in the power spectra of the harmonic coefficients calculated from: *a*, the Kitt Peak data (using the full latitude resolution); *b*, the Kitt Peak data after spatial degrading by averaging over the 30 bins in sine (latitude) used for the Mt Wilson data format; *c*, the Mt Wilson data. The boundary of the region ($l \leq 14$) free from spatial-resolution effects (when the Mt Wilson format is used) is indicated by a long tick mark. *A*, Odd values of l ; *B*, even values.

however, the effects are quite large, particularly in the odd-parity diagrams.

Next let us compare the Mt Wilson data with the Kitt Peak data, spatially degraded to the Mt Wilson format. The results are consistent with each other in the following sense. (1) Both data sets show that the odd- and even-parity modes form two decoupled systems. (2) Since the Mt Wilson time series (17.2 yr) is about twice as long as that of Kitt Peak (9.6 yr), the power estimates of the Kitt Peak data are much less stable. (3) Even though the Kitt Peak time series is much shorter than the 22-yr period, this resonance of the odd-parity modes is revealed in the Kitt Peak data, although the power tends to be shifted to higher frequencies. This shift is, however, entirely a consequence of the shortness of the time series. (4) Within the statistical fluctuations, the Kitt Peak and Mt Wilson data sets separately both show the ν - l behaviour for the even-parity modes given by Fig. 2. (5) In addition, the Kitt Peak data for the even-parity modes show some higher-frequency peaks, the statistical significance of which is uncertain.

The comparisons that we have made in Fig. 3 show that we can have confidence in the general results of Fig. 2, but that the length of the time series is critical to obtaining statistically stable results. This stability is more important than the frequency resolution provided by a long time series. The combined series of 25 yr seems to give us the needed stability.

It should be possible to calculate the harmonic coefficients up to an l value equal to half the number of sine (latitude) bins in the synoptic chart minus one, without running into spatial resolution problems. As the Kitt Peak synoptic maps contain 180 sine (latitude) bins between the north and south heliographic poles, it should thus be useful to consider modes up to $l = 89$ in the case of the Kitt Peak data. The results of such calculations are shown in Fig. 4.

For Fig. 4, the combined Mt Wilson-Kitt Peak data set has been used for $l \leq 14$, the Kitt Peak data alone for larger values of l . The resolution for the Kitt Peak data is shown by the open box to the right. The frequency resolution for the combined Mt Wilson-Kitt Peak data is better by almost a factor of three, which is also evident from the sharpness of the power peaks for $l \leq 14$.

The difference in statistical stability between the regions with l below and above 14 is also striking. In the region of the combined data set, not only is the frequency location of the resonances more well defined, but the noise in the form of the

background 'continuous' spectrum is much reduced. The dependence of the background on the length of the time series suggests that the intrinsic solar continuous background is weak, and that the 'true' spectrum may be a typical 'emission-line' spectrum.

In spite of the considerable stochastic noise fluctuations in Fig. 4, some general conclusions are possible. (1) The ν - l pattern contains discrete bands all the way up to $l = 89$. With increased resolution and stability, these bands might well narrow down to sharply defined resonances of large coherence length. (2) The diagrams of odd and even parity tend to be more similar to each other for $l > 15$ than in the case of Fig. 2 (for $l \leq 14$), although many apparently significant differences occur. (3) The 22-yr cycle is absent for all modes with $l \geq 15$ (see also Fig. 3), but is dominant for smaller odd l values. The 22-yr resonance

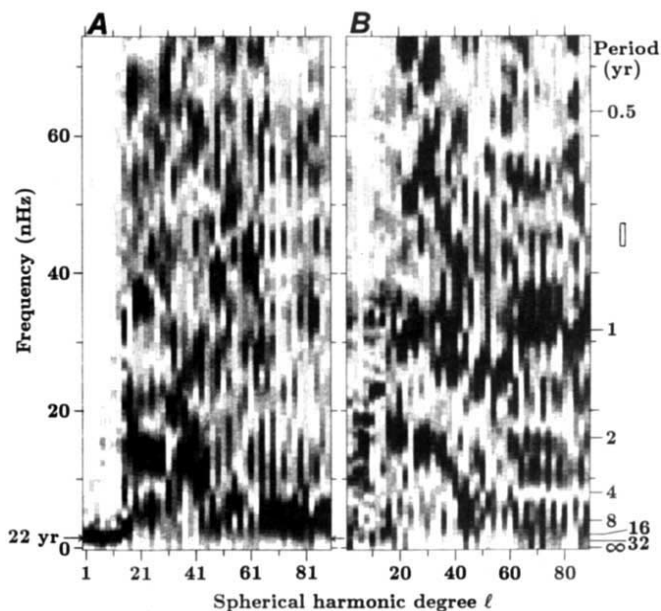


Fig. 4 Power spectra of the harmonic coefficients for $l \leq 89$. The combined Mt Wilson-Kitt Peak data set has been used for $l \leq 14$, the Kitt Peak data set alone for larger values of l . The resolution is indicated by the open box to the right. *A*, Odd values of l ; *B*, even values.

may be related to the parity symmetry-breaking mechanism, since the decoupling between odd and even parity appears to be complete only where the 22-yr cycle is present. The qualitative change occurring at $l \approx 15$ has nothing to do with the choice of the Mt Wilson data format of 30 sine (latitude) bins, as it is seen independently in the Kitt Peak data alone (Fig. 3), which have a much higher resolution (180 bins). (4) There is substantial power for odd l at periods of 2–3 yr when $17 \leq l \leq 39$, and around 6–9 yr for $l \geq 65$, which warns us not to carry the analogy with p-mode oscillations too far.

Only with an extended Kitt Peak time series can the noise be sufficiently suppressed to allow more definite quantitative statements to be made. The presence of structures at all l values up to 89 emphasizes the need for a very long continuous database of synoptic magnetic-field observations of highest possible spatial resolution, as carried out in the Kitt Peak programme. At the same time, it is important that the Mt Wilson programme should also continue, since these synoptic observations started 17 years before the corresponding Kitt Peak programme, and thus provide the best information for modes with $l < 15$.

Discussion

The present analysis has revealed a complex resonant modal structure in the evolution of the pattern of solar magnetic fields. The coherence length of these resonances appears to be greater than the length of the time series used, which is 25 yr for the combined Mt Wilson–Kitt Peak data. A parity selection rule, which is particularly striking for spherical harmonic degree $l < 15$, determines the pattern in ν – l space. Thus, for $l < 15$, the 22-yr cycle is the only feature of the odd-parity modes, whereas the modes of even parity show a ν – l dependence reminiscent of global standing waves in the solar interior. For $l \geq 15$ the parity symmetry breaking is a much less dominant feature, and the 22-yr periodicity is absent. The complex resonant pattern in ν – l space continues until the highest-order modes resolved, with $l = 89$.

The next question that arises concerns the nature of the discrete modes of even parity for $l \leq 14$. The periods of years are clearly too long for the pressure or the gravity force to have a dominant role. As the resonances occur in the magnetic-field pattern, it is natural to imagine that the electromagnetic force plays a central role. Let us now examine whether an interpretation in terms of internal waves makes sense.

The relation between angular frequency ω , wavenumber κ , and wave speed c is

$$\omega = kc \quad (4)$$

In analogy with the treatment of p-mode oscillations¹⁹, we assume that the wavenumber vector becomes horizontal at some radial distance r_i from the centre of the Sun, so that

$$k = \sqrt{l(l+1)}/r_i \quad (5)$$

there. The nature of the waves is not known, but possible candidates are dynamo waves and hydromagnetic waves. For any mechanism involving hydromagnetic waves it is natural to identify c with the speed of Alfvén waves (in a compressible medium hydromagnetic waves will not be pure Alfvén waves, but the use of the Alfvén speed is appropriate for our purpose

of obtaining an order-of-magnitude estimate):

$$c = B/\sqrt{4\pi\rho} \quad (6)$$

B being the magnetic field strength and ρ the gas density. We further use

$$\omega = 2\pi/P \quad (7)$$

P being the wave period. Then, we obtain an expression for the field strength required:

$$B = 4\pi^{3/2} \frac{r_i \sqrt{\rho}}{P \sqrt{l(l+1)}} \quad (8)$$

As we have no developed theory for these waves, and as we do not know the internal magnetic-field distribution in the Sun, we do not know at what distance r_i to apply equation (8). Thus, we will only select some values representative of the convection zone to indicate the order of magnitude of the magnetic field required. Let us as an example take the magnetic mode with $l = 6$, which according to Fig. 2 has a period P of 2 yr. Inserting this in equation (8), and taking ρ from a convection-zone model²⁰, we obtain $B = 1,200$ G if the data for the bottom of the convection zone are used ($r_i/R_\odot = 0.72$), whereas $B = 740$ G at the middle of the convection zone ($r_i/R_\odot = 0.86$). These values of B seem reasonable, because they are similar to the values we observe directly on the solar surface, in sunspots, faculae, or the quiet network. Our estimates and method of calculation are similar to that used to determine the possible torsional oscillations of the Sun²¹.

Most previous attempts to explain the evolution of solar magnetic fields have almost exclusively been made in terms of dynamo theory^{7–9}. It seems doubtful, however, that current dynamo theories provide a good theoretical framework for the explanation of the resonant structure of the even-parity modes. Although the kinematic dynamo equations decouple for the odd- and even-parity modes^{22,23}, both categories of modes are excited with about equal ease, and no variation of cycle frequency with l number or parity is predicted.

The symmetry breaking between odd and even parity is, however, entirely consistent with Hale's polarity law, which states that the orientation of the plus and minus magnetic polarities in sunspot groups is opposite in the two hemispheres, and reverses sign with the beginning of a new 11-yr sunspot cycle. This rule is strictly obeyed on the Sun and implies that the toroidal flux system giving rise to the sunspots is purely antisymmetric with respect to the equator (odd parity), and varies with a 22-yr cycle. A recent attempt to explain this polarity law in terms of dynamo theory has been made²⁴, invoking anisotropic diffusivity to generate a difference in growth rates between odd and even modes.

We thank Gordon Newkirk for providing us with the Mt Wilson data compiled and digitized in the 'Boulder format' (30×36 matrix per synoptic chart, with data on corrections for sunspot saturation) at the High Altitude Observatory. The Mt Wilson synoptic programme was set up and directed by Robert Howard. We also thank J. W. Harvey for providing the full Kitt Peak data set of high-resolution synoptic maps of the magnetic field.

Received 14 July; accepted 14 October 1985.

- Duvall, T. R. Jr *Trans. IAU* (Rep. IAU Commission 12, in the press).
- Deubner, F. L. & Gough, D. A. *Rev. Astr. Astrophys.* **22**, 593–619 (1984).
- Delache, P. & Scherrer, P. H. *Nature* **306**, 651–653 (1983).
- Scherrer, P. H. in *Solar Seismology from Space* (eds Ulrich, R. K. et al.) 173–182 (JPL Publ. 84–84, Pasadena, 1984).
- Fröhlich, C. & Delache, P. in *Solar Seismology from Space* (eds Ulrich, R. K. et al.) 183–193 (JPL Publ. 84–84, Pasadena, 1984).
- Stenflo, J. O. *Proc. 3rd Eur. Solar Meet.*, Oxford 11–34 (1981).
- Schüssler, M. *IAU Symp.* No. 102, 213–236 (1983).
- Gilman, P. A. *IAU Symp.* No. 102, 247–270 (1983).
- Yoshimura, H. *Astrophys. J. Suppl.* **52**, 363–385 (1983).
- Altschuler, M. D., Trotter, D. E., Newkirk, G. N. Jr & Howard, R. *Sol. Phys.* **39**, 3–17 (1974).
- Stenflo, J. O. *Sol. Phys.* **23**, 307–339 (1972).

- Altschuler, M. D. & Newkirk, G. N. Jr *Sol. Phys.* **9**, 131–149 (1969).
- Howard, R. & Stenflo, J. O. *Sol. Phys.* **22**, 402–417 (1972).
- Levine, R. H. *Sol. Phys.* **54**, 327–341 (1977).
- Altschuler, M. D., Levine, R. H., Stix, M. & Harvey, J. *Sol. Phys.* **51**, 345–375 (1977).
- Hoeksema, J. T. thesis. Stanford Univ. (1984).
- Howard, R., Tanenbaum, A. S. & Wilcox, J. M. *Sol. Phys.* **4**, 286–299 (1968).
- Howard, R., Boyden, J. E., Bruning, D. H., Clark, M. K., Christ, H. W. & Labonte, B. J. *Sol. Phys.* **87**, 195–203 (1983).
- Christensen-Dalsgaard, J., Duvall, T. L. Jr, Gough, D. O., Harvey, J. W. & Rhodes, E. J. *Jr Nature* **315**, 378–382 (1985).
- Spruit, H. C. *Sol. Phys.* **34**, 277–290 (1974).
- Layzer, D., Rosner, R. & Doyle, H. T. *Astrophys. J.* **229**, 1126–1137 (1979).
- Steenbeck, M. & Krause, F. *Astr. Nachr.* **291**, 49–84 (1969).
- Deinzer, W. & Stix, M. *Astr. Astrophys.* **12**, 111–119 (1971).
- Yoshimura, H., Wu, F. & Wang, Z. *Astrophys. J.* **285**, 325–338 (1984).

A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes

Francois Rouyer, Marie-Christine Simmler, Christophe Johnsson, Gilles Vergnaud, Howard J. Cooke* & Jean Weissenbach

Unité de Recombinaison et Expression Génétique, INSERM U163, CNRS UA271, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex, France

* MRC Mammalian Genome Unit, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK

Three independent pseudoautosomal loci are linked to sex determination at frequencies which define a gradient of linkage. The segregation patterns of these loci indicate that X/Y recombination results from a single obligatory meiotic crossing-over in the pseudoautosomal region. Recombination in male germ cells in the terminal regions of the short arms of the X and Y chromosomes is 10-fold greater than between the same regions of the X chromosomes in female germ cells.

THE distal parts of the short arms of the human X and Y chromosomes contain homologous sequences^{1,2} and have identical telomeres.³ Alleles of some of these homologous sequences undergo reciprocal interchange between the X and Y chromosomes^{1,3}. Thus, the distal short arms of the human sex chromosomes probably consist of a pseudoautosomal region⁴ characterized by the genetic identity⁵ of loci undergoing recombination. To test whether these exchanges result from gene conversion or cross-over, we followed segregation of three independent polymorphic pseudoautosomal DNA loci in eight large families. Results of this analysis give a picture of the genetic properties of the human X/Y pseudoautosomal region.

Extensive variability

Two of the pseudoautosomal loci studied have been described previously^{1,3}. Hybridization patterns of probe 29C1 (ref. 3), which detects an extremely polymorphic DNA locus, *DXYS14* (refs 6, 7), located close to the telomere of the sex chromosomes, and of probe 113D, which defines another highly polymorphic pseudoautosomal DNA locus, *DXYS15* (refs 1, 6, 7), are shown in Figs 1 and 2. The third probe (601) used in this study originates from a complementary DNA library (see Fig. 1 legend) and defines an additional pseudoautosomal locus, *DXYS17* (refs 6, 7), mapping to Xp22.3 and Yp. Probe 601 detects many restriction-fragment length polymorphisms (RFLPs) in different restriction enzymes digests. Not all these RFLPs co-segregate and their molecular basis is not yet fully determined. Two of them are illustrated in Figs 1c, d and 2c.

The *DXYS17* *TaqI* RFLP consists of a series of at least eight allelic fragments ranging from 1.9 to 1.1 kilobases (kb) (alleles *DXYS17 T1-DXYS17 T8*). The size variations of the *TaqI* RFLP can be observed with other restriction digests and may therefore be ascribed to an insertion/deletion process. The segregation of three allelic *TaqI* fragments (*T3*, *T4* and *T6*) is shown in Fig. 1d. The *DXYS17* *EcoRI* RFLP (Fig. 1c) consists of two allelic fragments of 17 kb (*DXYS17 R1*) and ~7 kb (*DXYS17 R2*) which occur in the population with a frequency of 30 and 70%, respectively. This size difference has not been found in the other RFLPs detected by 601, indicating that the *EcoRI* RFLP probably originated from a point mutation with loss or gain of *EcoRI* site. In addition, *DXYS17 R2* is stable in a single family, but can fluctuate among unrelated individuals from ~6.5 to 7.5 kb. In some of the individuals homozygous for the *DXYS17 R2* allele, it is possible to distinguish the two forms of this allele. This is consistent with the superimposition of a second type of variation possibly related to the *DXYS17* *TaqI* polymorphism. Probably the same variations occur in the *DXYS17 R1* allele, but without detectable effect on its electrophoretic mobility. Both the *DXYS17* *EcoRI* and *TaqI* alleles are inherited in a Mendelian fashion and their combinations define several haplotypes (see Fig. 1 legend).

TaqI and *EcoRI* DNA digests of 12 unrelated individuals have been hybridized to probe 601 (not shown) and 11 different haplotypes can be deduced from the different allelic combinations. Strikingly, the *DXYS17 R1* and *DXYS17 R2* alleles can both be associated with alleles *DXYS17 T3*, *DXYS17 T5*, *DXYS17 T6* and *DXYS17 T8*. Assuming that a single allele can only exceptionally arise twice independently, it is highly unlikely that the two *DXYS17 R* alleles can be associated with more than one common *DXYS17 T* allele. This strongly suggests that several recombination events have occurred between both polymorphic sites and have generated new haplotypes in which both *DXYS17 R* alleles have been reassorted with new *DXYS17 T* alleles.

Thus, all three probes used in this study detect highly polymorphic DNA sequences. Similarly, pseudoautosomal loci such as *DXYS20* (refs 6, 7) and others (G.V., unpublished data) also exhibit highly polymorphic restriction fragments. In general, isolation of probes detecting highly polymorphic sequences is rare. As only very few additional pseudoautosomal probes have been reported¹, this region appears to be very diverse. With the exception of the *DXYS17* *EcoRI* polymorphism, RFLPs associated to these three probes have not been ascribed to point mutations. It is not clear yet whether these polymorphisms have originated through unequal recombination events, as proposed recently for minisatellite sequences⁸, or by some other mechanism.

Recombination of pseudoautosomal loci

To determine recombination frequency, DNA samples from large families were analysed for the RFLPs detected by pseudoautosomal probes 29C1, 113D and 601. *TaqI* or *EcoRI* Southern blots of the available samples of the paternal grandparents, parents and children from eight families were probed with all three pseudoautosomal sequences. As a control, the sex of each individual has been reconfirmed using probe 49f which detects multiple Yq-specific *TaqI* fragments, some of which are polymorphic. Many haplotypes of the human Y chromosome can be defined from the different existing combinations of the polymorphic 49f *TaqI* fragments⁹. All males of a family showed the same 49f haplotype (data not shown). Segregation of alleles from the different pseudoautosomal loci in paternal and maternal meiosis is analysed for two families in Figs 1 and 2 legends. Recombination for all three loci occurs in at least one of these families, as shown in Figs 1 and 2. Segregation of *DXYS14*, *DXYS15* and *DXYS17* with respect to sexual phenotype has been analysed for the eight families and recombination events are indicated in Table 1. The overall results obtained for segregation between sex and loci *DXYS14*, *DXYS15* and *DXYS17* are also summarized in Table 1. These family studies show that three pseudoautosomal loci, detected by randomly derived DNA probes, can be transferred from one sex chromosome to the

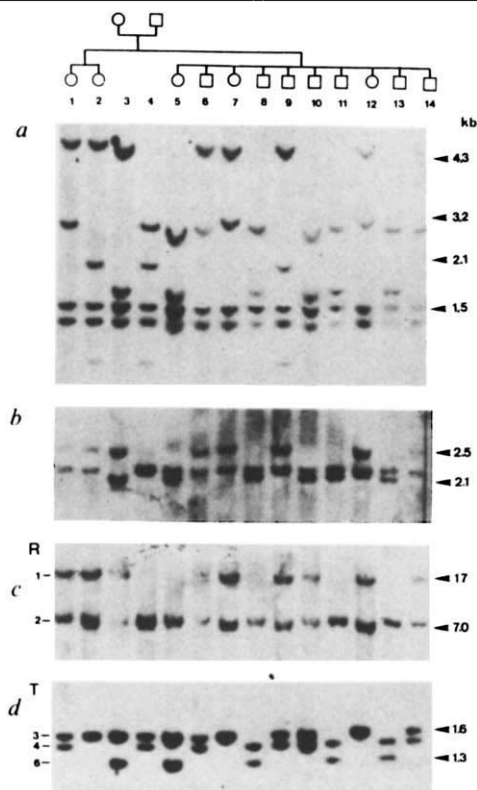


Fig. 1 Pedigree of family PO and hybridization patterns to Southern blots of restriction digests of DNAs from the family. Circles, females; squares, males. *a*, *TaqI* digests probed with 29Cl (ref. 3); *b*, *TaqI* digests probed with 113D (ref. 1); *c*, *d*, *EcoRI* and *TaqI* digests, respectively, probed with 601.

Methods. DNA (5–12 µg) was completely digested with an excess of restriction endonuclease, electrophoresed on a 1% agarose gel, transferred to nylon membranes and irradiated with a standard ultraviolet transilluminator for 5 min. After prehybridization at 42 °C for 4 h in 50% formamide, 5×SSX (1×SSC: 0.15 M NaCl, 15 mM Na-citrate pH 7), 5×Denhardt's solution (Denhardt's solution: 0.02% Ficoll 400, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 20 mM Na-phosphate pH 6.6 and 50 µg ml⁻¹ denatured sonicated salmon-sperm DNA, filters were hybridized overnight at 42 °C in the same solution with addition of 10% dextran sulphate and 10–30 × 10⁶ c.p.m. of ³²P-labelled DNA probe. Probes used were DNA inserts nick-translated to a specific activity >10⁸ c.p.m. µg⁻¹. After washing, the filters were autoradiographed with Kodak XAR5 films backed with one or two intensifying screens for 1–4 days at –70 °C. Probe 601 is isolated from a cDNA library derived from the 3E7 hybrid cell line (C.J., manuscript in preparation), a mouse-human somatic hybrid containing Y as the only recognized human chromosome. After hybridization, the filters were washed in 2×SSC. Segregation analysis: definition of haplotypes: In this family, the higher maternal *TaqI* *DXYS17* fragment of 1.6 kb (*DXYS17* T3) (*d*) and *EcoRI* *DXYS17* R1 allele (*c*) are strictly linked and hence define a first haplotype (*DXYS17* T3–R1), whereas the second maternal haplotype (*DXYS17* T6–R2) consists of the lower 1.3-kb *TaqI* allele (*DXYS17* T6) associated with the lower *EcoRI* fragment *DXYS17* R2. Combinations of paternal alleles define two other haplotypes which consist of the lower *EcoRI* *DXYS17* R2 allele associated either with the higher *TaqI* *DXYS17* T3 fragment (haplotype *DXYS17* T3–R2) or with the intermediate one of 1.5 kb (*DXYS17* T4) (haplotype *DXYS17* T4–R2). Maternal meiosis: Maternal heterozygosity for all four RFLPs (*DXYS14*, *DXYS15*, *DXYS17* T and *DXYS17* R) and a difference among parents for at least one allele, makes family PO informative for recombination in maternal meiosis. Maternal haplotype *DXYS17* T3–R1 segregates with the higher *DXYS15* allelic fragment of 2.5 kb in cases 1, 2, 6, 7, 9, 12 and 14 and haplotype *DXYS17* T6–R2 with the lower *DXYS15* allelic fragment of 2.1 kb in cases 5, 8, 11 and 13. This association is not observed in case 10 where the lower *DXYS15* allele co-segregates with haplotype *DXYS17* T3–R1. The maternal X chromosome of individual 10 therefore results from a recombination between *DXYS15* and *DXYS17*. Paternal meiosis: The phase of paternal alleles of pseudoautosomal loci of the family has been determined from a study of the grandparents (not shown). The *DXYS14* paternal Y allele is found in daughters 1, 5, 7 and 12. Conversely the paternal X allele is found in son 9. These five children thus carry a recombinant paternal sex chromosome and the seven remaining children carry the same *DXYS14* allele as their father on their paternal X or Y chromosome. As the father is homozygous for locus *DXYS15* the family is not informative for segregation of this locus. All seven sons inherit paternal Y haplotype *DXYS17* R2–T4, whereas one daughter (number 1) out of five inherits this haplotype as well and is the only child showing recombination for *DXYS17*.

other. Assuming that the interchange event is strictly reciprocal, recombination should occur in both male and female progeny with the same probability. For *DXYS14* where there are sufficient data, this expectation is fulfilled.

According to the segregation analysis of family PO (Fig. 1), the maternal X chromosome of individual 10 results from a recombination between *DXYS15* and *DXYS17*, indicating that these loci are not tightly linked. A similar conclusion can be drawn for *DXYS14*, which recombines with *DXYS17* in this individual. The recombination data for the pseudoautosomal loci in female meiosis are given in Table 2. No recombination between *DXYS14* and *DXYS15* has been observed in female meiosis.

A single obligatory X–Y crossover

All the families presented in Table 1 are informative for three-point linkage analysis (sexual phenotype and two pseudoautosomal loci) (see Table 3 for summary). The assumption that each pseudoautosomal locus may reassort independently in male meiosis predicts a frequency for each of the different possible combinations, but the less frequently recombined locus *DXYS17* never segregated independently from either *DXYS14* or *DXYS15* (Tables 1, 3). Similarly, a recombination of locus *DXYS15* always involved the most frequently recombined locus *DXYS14* (Table 1, 3). In addition, some of the frequencies predicted if in equilibrium are in total discrepancy with the observed values: in type *a* crosses (Table 3) we would expect, respectively, 16% of double recombinants (0.5×0.315), 34% of single recombinants with *DXYS14* [$0.5 \times (1 - 0.315)$], 16% of single recombinants with *DXYS15* [$(1 - 0.5) \times 0.315$] and 34% of non-recombined paternal sex chromosomes [$(1 - 0.5) \times (1 - 0.315)$]. However, we observe 34% of double recombinants, 14% of recombinants mobilizing *DXYS14* exclusively, no recombinants with *DXYS15* exclusively and 51% non-recombinants. In type *c* crosses (Table 3) an independent reassortment of pseudoautosomal loci in male meiosis also predicts, that *DXYS17* co-segregates with *DXYS15* in 31% of the *DXYS17* recombinants, whereas in the other 69% *DXYS17* should recombine without *DXYS15*. Instead we observe co-segregation of both loci in all *DXYS17* recombinants (Table 3). These three-point analyses eliminate the possibility of an independent reassortment of pseudoautosomal loci in male meiosis. Recombination frequencies between pseudoautosomal loci range between 14.5 and 36% (Table 2), indicating an obvious linkage between them. Co-segregation of three independent pseudoautosomal loci is not consistent with recombination via gene conversion which is supposed to take place on short stretches of DNA and should result in an independent segregation of each locus. On the contrary, linkage of the three loci and absence of independent segregation of the less frequently recombining ones demonstrate the occurrence of crossing-over events in the pseudoautosomal regions of the X and Y chromosomes.

As inferred from Table 1, pseudoautosomal loci display either partial linkage to their paternal sex chromosomes with recombination rates <50% for *DXYS15* and *DXYS17* or no sex linkage with a recombination frequency of ~50% for *DXYS14*. As we have established that sex chromosomes recombine through crossing-over, partial sex linkage can be interpreted in several ways: (1) X/Y crossing-over occurs occasionally during meiosis; (2) an obligatory X/Y crossing-over takes place at different locations from one meiosis to another and thus does not always mobilize the same loci; and (3) a few X/Y crossovers occur during a single meiosis. If X/Y crossing-over is occasional, frequencies of 50% would not be observed with the telomeric locus *DXYS14*. Nor are frequencies of 50% consistent with the occurrence of several crossovers involving the same chromatids during a single meiosis. In addition, if a single X/Y pair underwent multiple recombinations, there should be examples of independent segregation of each of the studied loci. The two

Table 1 Segregation of loci *DXYS14*, *DXYS15* and *DXYS17* with respect to sexual phenotype in male meiosis

Family	Proband	Sex	Recombination with <i>DXYS14</i>	Recombination with <i>DXYS15</i>	Recombination with <i>DXYS17</i>	Family	Proband	Sex	Recombination with <i>DXYS14</i>	Recombination with <i>DXYS15</i>	Recombination with <i>DXYS17</i>	
NG	42	M	+	NI	—	K1413	1085	M	+	NI	—	
	35	F	—	NI	—		1084	M	+	NI	+	
	36	F	+	NI	—		1072	M	+	NI	+	
	37	M	—	NI	—		1092	F	+	NI	—	
	43	F	+	NI	+		1091	M	—	NI	—	
	26	M	—	NI	—		1279	M	—	NI	—	
PO	01	F	+	NI	+	1081	M	—	NI	—		
	02	F	—	NI	—	1078	M	—	NI	—		
	05	F	+	NI	—	1079	F	+	NI	—		
	06	M	—	NI	—	1080	M	—	NI	—		
	07	F	+	NI	—	1076	M	+	NI	—		
	08	M	—	NI	—	1087	M	+	NI	—		
	09	M	+	NI	—	1088	F	+	NI	—		
	10	M	—	NI	—	1090	M	+	NI	—		
	11	M	—	NI	—	K1418	1262	F	—	—	NI	
	11	F	+	NI	—		1267	F	+	+	NI	
	13	M	—	NI	—		1260	F	+	—	NI	
	14	M	—	NI	—		1261	F	+	—	NI	
	66	6603	M	+	+		—	1256	F	—	—	NI
		6609	M	+	+		—	1263	F	—	—	NI
6606		M	—	—	—		1258	M	+	+	NI	
6607		F	—	—	—		1259	M	+	+	NI	
6604		F	—	—	—	K1421	1309	F	+	+	—	
K1332		8252	F	+	+		NI	1326	F	+	+	+
	8265	F	—	—	NI		1308	F	—	—	—	
	8261	M	+	—	NI		1306	F	—	—	—	
	8256	F	—	—	NI		1307	F	+	+	+	
	8255	F	—	—	NI		1560	M	—	—	—	
	8259	F	—	—	NI		1543	M	—	—	—	
	8253	M	+	—	NI		1304	M	+	+	—	
	8254	M	—	—	NI	Total recombined X chromosomes			15	5	4	
K1346	8364	F	—	—	—	Total recombined Y chromosomes			18	6	3	
	8362	M	—	—	ND	Total recombined sex chromosomes			33	11	7	
	8357	M	+	+	+	Total meioses with X			30	19	19	
	8603	M	—	—	—	Total meioses with Y			37	16	31	
	8359	M	+	—	—	Total meioses			67	35	50	
	1370	M	—	—	—	Recombined X chromosomes (%)			50	26.5	21	
							Recombined Y chromosomes (%)			48.5	37.5	9.5
							Recombined sex chromosomes (%)			49.5	31.5	14

(+), Involvement of a pseudoautosomal locus in a recombination affecting a paternal sex chromosome; (—), absence of recombination between a pseudoautosomal locus and sexual phenotype; NI, non-informative paternal phenotype; ND, not determined. Segregation of pseudoautosomal loci has been analysed on Southern blots with probe 29C1 for locus *DXYS14*, probe 113D for locus *DXYS15* and probe 601 for locus *DXYS17*. DNA from each individual was digested with *TaqI* and with *EcoRI* for DNAs from families PO, K1346 and K1421. Chromosomal assignment for the paternal alleles is deduced from an analysis of the paternal grandparents. DNA samples from families 28, 66, K1332, K1346, K1413, K1418 and K1421 were provided by CEPH. Family 28 was only informative for female meiosis and is not shown here.

most widely separated loci, *DXYS14* and *DXYS17*, are 36 recombination units apart. In *DXYS17* recombinants, a second crossing-over proximal to the telomeric locus *DXYS14* is theoretically possible in such a recombination interval. No example of double recombination has been found in any of the seven *DXYS17* recombinants analysed here (Table 3). Similarly, no second event is found in the smaller *DXYS14/DXYS15* interval (14.5 units) in the 12 *DXYS15* recombinants. However, the small size of the sampling may be at the limit of statistical significance. Nonetheless a frequency of 50% for the telomere would otherwise require during each meiosis the occurrence of many recombination events which should be observed even in small samples. Furthermore, as shown below, in the pseudoauto-

somal region recombination frequencies can be added up to 50% without any distortion, again reflecting the existence of a single crossover event. Therefore, taken together, our results strongly support hypothesis (2) in which a single obligatory crossover accounts for a recombination frequency of 50% of the telomeric region and for co-segregation of the more distal loci in recombinations affecting the proximal ones.

Recombinational hotspot

The extent of the pseudoautosomal region remains unknown. According to cytogenetic measurements, the size of the Y chromosomal short arm is generally estimated to range from 5,000 to 10,000 kb whereas the size of the pairing region is

Table 2 Recombination between pseudoautosomal loci

Recombination between	<i>DXYS14/DXYS15</i>	<i>DXYS15/DXYS17</i>	<i>DXYS14/DXYS17</i>
Female meiosis			
Recombined loci	0	2	2
Meioses	48	40	52
Recombination rate (%)	0	5	4
Male meiosis			
Recombined loci	5	4	18
Meioses	35	18	50
Recombination rate (%)	14.5	22	36

Results for female meiosis are as follows: (1) Recombination between loci *DXYS14* and *DXYS15* was not observed in families PO, 28 (children 2803, 2804, 2805, 2806, 2807, 2808, 2809), 66 K1332, K1418 and K1421, other families being non-informative. (2) Recombination between loci *DXYS14* and *DXYS17* was observed in case 10 from family PO and in case 1304 from family K1421 whereas no recombination was observed in the rest of those two families as well as in families 28, 66, K1332 and K1346. The other families were not informative. (3) Recombination between loci *DXYS15* and *DXYS17* was again observed in case 10 from family PO and in case 1304 from family K1421, whereas the rest of these families and families 28, 66, and K1332 did not display recombination. Other families were not informative for these two loci. Results for male meiosis are derived from Table 1.

Table 3 Non-occurrence of double recombination events in male meiosis examined in three-point analyses

Three-point analysis	Type a	Type b	Type c
Pseudoautosomal loci involved in recombination			
<i>DXYS14</i> exclusively	5	18	—
<i>DXYS15</i> exclusively	0	—	4
<i>DXYS17</i> exclusively	—	0	0
Both pseudoautosomal loci	12	7	3
Nonrecombinants	18	25	11
Total meioses	35	50	18

Type a analysis considers segregation of *DXYS14*, *DXYS15* and sexual phenotype; b segregation of *DXYS14*, *DXYS17* and sexual phenotype; and c segregation of *DXYS15*, *DXYS17* and sexual phenotype.

variable and can even stretch beyond Yp (ref. 10). As many Yp-specific DNA sequences have no homologous counterparts on Xp (refs 11–13) and our unpublished results) it follows that X/Y synapsis is partly non-homologous and no accurate correlation between X/Y pairing and sequence homologies can be drawn. More than 90% of the Yp-specific DNA probes we have isolated to date are not pseudoautosomal (refs 11, 12 and our unpublished results). If our Yp-specific DNA probes and those obtained at random by the different laboratories are representative of the DNA composition of the chromosome short arm, it seems unlikely that the maximum size of the pseudoautosomal region is >5,000 kb.

The recombination frequency between *DXYS14* and *DXYS17* (Table 2) in male meiosis (36%) is ninefold higher than that obtained from female meiosis (4%) and contrasts with a significant general increase reported for the autosomal frequency in female meiosis as compared with that in male meiosis^{14,15}. In the human, 1% recombination (1 centimorgan) is generally taken to represent a chromosomal distance of 1,000 kb (ref. 16) and accordingly, loci *DXYS14* and *DXYS17* should be located 36,000 kb apart. The actual size of this interval is probably one order of magnitude lower as reflected by the recombination frequency observed in female meiosis (4%). It thus appears that the pseudoautosomal region is a recombinational hotspot in male meiosis. Such an overestimation is consistent with the view that a single crossover between the X and Y chromosomes is

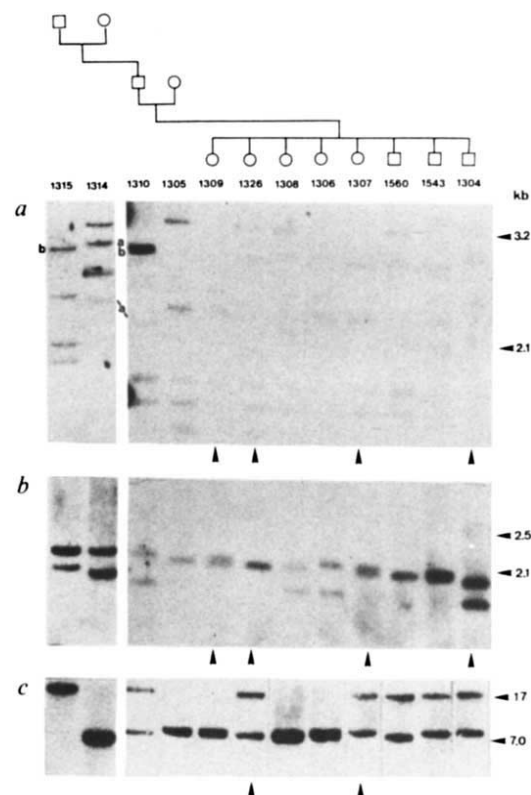


Fig. 2 Pedigree of CEPH family K1421 and hybridization patterns to Southern blots of restriction digests of the family. DNA digestions, blotting, hybridizations and washings were as described in Fig. 1. a, *TaqI* digestions probed with 29C1 (ref. 3); b, *TaqI* digestions probed with 113D (ref. 1); c, *EcoRI* digestions probed with 601. Segregation analysis: genotypes of grandparents have been determined independently and are displayed in the left part of the figure. In this family the father (1310) has inherited the *DXYS14* a allele (represented by two fragments) from his mother (1314), a is thus X-located whereas the b allele, originating from his father (1315), must be Y-located (a). Daughters 1309, 1326 and 1307 carry the paternal Y allele b. As the X-located paternal allele a is not found in these three daughters the paternal b allele must now be carried by their X chromosome. Conversely the paternal X allele a must be carried on the paternal Y chromosome in son 1304. There is no recombination of a or b alleles with sex phenotype in children 1308, 1306, 1560, 1543. Paternal X and Y alleles of locus *DXYS15* co-segregate strictly with paternal X and Y alleles of locus *DXYS14* (b). The paternal Y allele of locus *DXYS17* is only found in daughters 1326 and 1307 whereas daughter 1309 and son 1304, like the other children, show no recombination for this locus (c).

an obligatory step in male meiosis. It has been proposed that chiasmata play a part in the disjunction of chromosomes¹⁷ during the first meiotic division and that at least one chiasma takes place in each bivalent to ensure proper segregation. This view could apply to the pseudoautosomal region of the X/Y pair. The apparent 10 fold increase in recombination frequency in males compared with females is then a direct consequence of a chiasma having to be formed between each chromosomal pair in this limited region.

Genetic map

As the recombination frequency is extremely high in male meiosis, a detailed genetic map of the pseudoautosomal region can be constructed. Genetic distances between the pseudoautosomal loci are given in Table 2. According to these values *DXYS15* is located within the *DXYS14/DXYS17* interval (Fig. 3). Strikingly, the sum of the distances between *DXYS14* and *DXYS15* and between *DXYS15* and *DXYS17* is almost equal

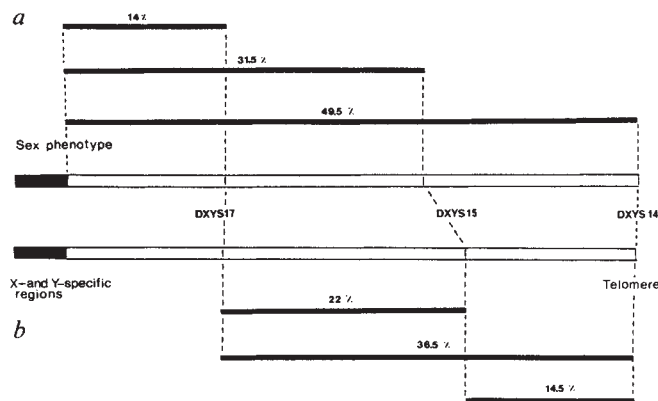


Fig. 3 Correlation of two genetic maps of the human pseudoautosomal region. *a*, Map derived from the recombination frequencies of pseudoautosomal loci with sexual phenotype; *b*, map derived from the recombination frequencies between pseudoautosomal loci in male meiosis.

to the value of the *DXYS14/DXYS17* interval. It appears that the recombination frequency additivity rule applies to the pseudoautosomal region for values of up to 36%. It may even apply to the entire pseudoautosomal region, as the distance between sex and *DXYS17* (14%) is again extremely close to the value obtained by subtracting the *DXYS14/DXYS17* distance from the distance between *DXYS14* and sex (49.5–36%). Despite a slight discrepancy between maps *a* and *b* in the position of *DXYS15* (Fig. 3) which remains to be investigated, these results demonstrate the occurrence of a single crossing-over in the pseudoautosomal region. Thus, a gradient of sex linkage can be correlated with the linear order of the pseudoautosomal loci. In other words, the recombination frequency directly reflects the position of the locus with respect to the regions specific to one of the sex chromosomes. The telomere is being exchanged at a rate of 50% whereas the more proximal loci are mobilized less frequently.

To date, construction of a deletion map of the human Y chromosome has been achieved using non-pseudoautosomal probes¹¹ and the testis-determining factor has been definitely assigned to Yp (refs 11, 18). Recombination mapping of the Y chromosome pseudoautosomal region represents a new method of approaching the sex determination region. Highly polymorphic pseudoautosomal probes also provide a supplementary tool for the analysis of the aetiology of XX maleness by allowing assessment of the paternal origin of the pseudoautosomal region.

The elevated recombination frequency in the human pseudoautosomal region provides a valuable opportunity to correlate experimental values and physical distances as chromosome walking can be performed over distances of several hundred kilobases of DNA. Alternatively, the physical distance between two loci may be amenable to determination by pulse-field electrophoresis, and a size estimation of the whole

pseudoautosomal region could be drawn from measurements of recombination frequency. The human pseudoautosomal region also presents an opportunity to investigate chromosomal recombination by both molecular and genetic methods. A molecular analysis of the highly polymorphic pseudoautosomal loci could reveal new structural features directly related to the crossing-over mechanism.

Conclusions

Several characteristics of pseudoautosomal sequences have been demonstrated in this report: (1) they can be exchanged between the X and Y chromosomes by a crossing-over mechanism; (2) they can exhibit partial sex linkage, the extent of the linkage being different for three independent pseudoautosomal loci; and (3) the pseudoautosomal region is characterized by a high recombinational activity and thus lends itself to extensive mapping. This mapping is even facilitated by occurrence of a single crossing-over per meiosis taking place at varying locations in the region.

Crossing-over between the mammalian sex chromosomes was proposed many years ago¹⁹. Recombination nodules have even been observed within the distal short arms of the human X/Y pair²⁰. The present data validate these earlier statements and confirm the view that the mammalian sex chromosomes originated from an ancestral pair of homologues. Indirect genetic evidence has shown the existence of crossing-over in mouse^{21–24}. However in this species, both the steroid sulphatase (*STS*) gene²⁴ and the sex-reversing (*Sxr*) mutation^{21–23} seem to recombine with a 50% frequency, whereas in humans frequencies specific for each pseudoautosomal locus have been observed. The *Sxr* mutation is probably distal to the mouse pseudoautosomal region^{22,23} and a frequency of 50% (ref. 24) should be expected. A frequency of 50% observed with the *STS* gene indicates that crossing-over is obligate for this locus either because it is very distal in mouse or because there is only one or a cluster of very few crossover points in the mouse X/Y pair. A small number of crossover points in the pseudoautosomal region indicates that this region is rather small in the mouse and contains only a few genes. This could explain why the lack of a set of these few genes in 39,X mouse is considerably less harmful²⁵ than the absence of one copy of the pseudoautosomal region in human 45,X Turner females²⁶.

Differences in genetic behaviour of pseudoautosomal regions between species may reflect a continuing process of sex-chromosome differentiation. Perhaps the size of the pseudoautosomal region is reduced by divergence of the most proximal areas which recombine at the lowest frequencies. We do not know if a reduction in recombination is a cause or a consequence of divergence, or if there is a discrete boundary to the pseudoautosomal region.

We thank the Centre d'Etude du Polymorphisme Humain (CEPH) for the family DNA samples; the CEPH staff for its kindness; Pierre Tiollais for encouragement; and Colin Bishop, Adrian Bird and William Brown for reading this manuscript. G.V. is a member of Délégation générale de l'Armement.

Received 23 September; accepted 12 November 1985.

1. Simmler, M. C. *et al.* *Nature* **317**, 692–697 (1985).
2. Buckle, V., Mondello, C., Darling, S., Craig, I. W. & Goodfellow, P. N. *Nature* **317**, 739–741 (1985).
3. Cooke, H. J., Brown, W. A. & Rappold, G. A. *Nature* **317**, 687–692 (1985).
4. Burgoyne, P. S. *Hum. Genet.* **61**, 85–90 (1982).
5. Polani, P. E. *Hum. Genet.* **60**, 207–211 (1982).
6. Willard, H. F. & Skolnick, M. H. *Cytogenet. Cell Genet.* (in the press).
7. Goodfellow, P. N., Davies, K. E. & Ropers, H. H. *Cytogenet. Cell Genet.* (in the press).
8. Jeffreys, A. C., Wilson, V. & Thein, S. L. *Nature* **314**, 67–73 (1985).
9. Ngo, K. Y., Vergnaud, G., Johnsson, C., Lucotte, G. & Weissenbach, J. *Am. J. Hum. Genet.* (in the press).
10. Chandley, A. C., Goetz, P., Hargreave, T. B., Joseph, A. M. & Speed, R. M. *Cytogenet. Cell Genet.* **38**, 241–247 (1984).
11. Vergnaud, G. *et al.* *Am. J. Hum. Genet.* (in the press).
12. Geldwerth, D. *et al.* *EMBO J.* **4**, 1739–1743 (1985).

13. Page, D. C., Harper, M. E., Love, J. & Botstein, D. *Nature* **311**, 119–123 (1984).
14. Cook, P. J. L. *Am. J. Hum. Genet.* **28**, 393–401 (1965).
15. Renwick, J. H. *Bull. Eur. Soc. Hum. Genet.* **2**, 7 (1968).
16. Renwick, J. H. *Br. med. Bull.* **25**, 65–73 (1969).
17. Darlington, C. D. *Recent Advances in Cytology* (Churchill, London, 1937).
18. Casanova, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (submitted).
19. Koller, P. C. & Darlington, C. D. *J. Genet.* **29**, 159–173 (1934).
20. Solari, A. J. *Chromosoma* **81**, 315–337 (1980).
21. Cattanaach, B. M., Pollard, C. E. & Hawkes, S. G. *Cytogenetics* **10**, 318–337 (1971).
22. Singh, L. & Jones, K. W. *Cell* **28**, 205–216 (1982).
23. Evans, E. P., Burtenshaw, M. D. & Cattanaach, B. M. *Nature* **300**, 443–445 (1982).
24. Keitges, E., Rivest, M., Siniscalco, M. & Gartler, S. M. *Nature* **315**, 226–227 (1985).
25. Russel, W. L., Brauch Ressel, L. & Gower, J. S. *Proc. Natn. Acad. Sci. U.S.A.* **45**, 544–560 (1959).
26. Ford, C. E., Jones, K. W., Polani, P. E., de Almeida, J. C. & Briggs, J. H. *Lancet* **i**, 711–713 (1959).

Large-scale dissociation of molecular gas in galaxies by newly formed stars

R. J. Allen^{*†}, P. D. Atherton^{*‡} & R. P. J. Tilanus^{*}

^{*} Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands

[†] Department of Astronomy, University of Illinois, 1011 West Springfield Avenue, Urbana, Illinois 61801, USA

[‡] Queensgate Instruments Ltd., 112 Windmill Road, Sudbury-on-Thames, Middlesex TW16 7HB, UK

Molecular hydrogen is thought to form from atomic hydrogen on the surface of dust grains in the interstellar medium, and to be dissociated primarily by ultraviolet radiation from hot stars¹. This process has been modelled on the 1 pc length scale for the case of dense interstellar clouds² and near newly formed ionizing stars³. Observational evidence that molecular hydrogen is indeed dissociated in shells around young OB stars in the Galaxy has been presented recently⁴⁻⁶. We suggest here that the same dissociation process occurs on the kpc scale in active star-forming regions of some galaxies, and that this dissociation may strongly affect the observed morphology of atomic hydrogen in spiral arms.

The large-scale distribution of molecular gas in galaxies is poorly known, not only because of the limited availability of observable tracers and the low angular resolution and sensitivity of the instrumentation but also because of the difficulties of interpreting the observed surface brightness of molecular lines as column densities of H₂. Observations of the CO molecule in several galaxies do, however, indicate that the general distributions of molecular and atomic gas appear to be different. In particular, the central H I depressions observed at radii ≤ 5 kpc in several prominent bright late-type spiral galaxies may be filled in with molecular gas^{7,8}. One such galaxy is M83 (NGC5236) and, although the angular resolution of the CO observations which are available at millimetre wavelengths is insufficient to discern details on the scale of a spiral arm, the ¹²CO is concentrated in the central regions within a radius of ~ 1 arc min (ref. 9). Recent observations of several galaxies in ¹²CO and ¹³CO with a resolution of 65 arc s confirm that molecular gas is present over the whole disk of M83, although the quantitative interpretation in terms of H₂ column density is uncertain¹⁰.

More detailed pictures of nearby galaxies can be made of the distribution of atomic H I using centimetre-wavelength aperture synthesis radio telescopes. Contrary to the situation with the molecular gas, the conversion of an observed surface brightness in the 21-cm line to a column density of H I atoms is straightforward and independent of the excitation temperature, provided the latter is substantially higher than the temperature of the cosmic microwave background radiation. Over a wide range of conditions relevant to the present case, the excitation temperature of H I is equal to the kinetic temperature of the interstellar gas²¹, and this is much higher than 3 K. It is highly unlikely that substantial quantities of H I remain undetected in the nearby galaxies.

We have observed M83 in H I with a maximum resolution of 10 arc s using the Very Large Array (VLA), and combined the data with lower resolution VLA observations by M. Ondrechen and J. M. van der Hulst (personal communication) to produce the final maps. Figure 1 shows the distribution of H I at 25 arc s resolution after subtraction of the continuum, along with a sketch of several prominent dust lanes. The warped outer H I 'ring' described by Rogstad *et al.*¹⁴ lies outside the frame of Fig. 1. Although the contour map does not show it very clearly, other methods of displaying this data indicate that the central region with diameter 2 arc min is essentially devoid of H I. In the following discussion, we argue that, independent of the

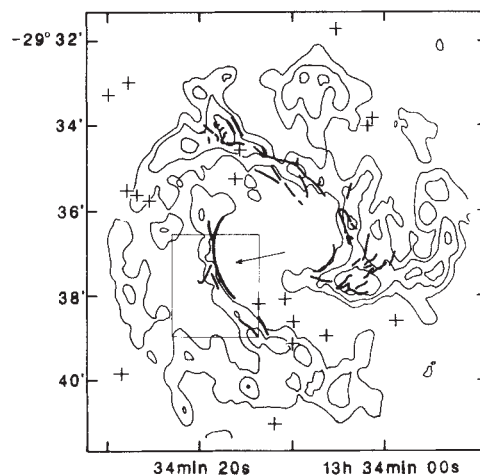


Fig. 1 The distribution of atomic hydrogen and dust lanes in M83. Contours are taken from VLA-H I synthesis observations at 25 arc s resolution. The dust lanes have been drawn from inspection of an optical photograph. The crosses refer to a set of secondary standard stars. The arrow indicates the inner eastern arm; the frame drawn corresponds approximately with the frames of Fig. 2.

controversy surrounding the interpretation of the CO observations, it is likely that the H I void in M83 is filled with molecular gas in quantities similar to the H I found at slightly larger radii.

As part of a project to study the temporal sequence of star formation in spiral arms, we have been examining in detail a portion of M83 which includes the boundary zone at the edge of the H I void. This region is indicated by the small frame and the arrow in Fig. 1. As a tracer for the young stars, we have mapped the galaxy in the optical H β emission line using the Taurus imaging Fabry-Pérot spectrometer¹¹ attached to the Anglo-Australian Telescope. In the region indicated by the small frame in Fig. 1, the distribution of H β emission is shown in Fig. 2a after subtraction of the continuum and smoothing to a resolution of 5 arc s. For comparison, Fig. 2b shows the H I distribution with a resolution of 10 arc s. The prominent dust lanes are indicated by the thick lines and the dashed line shows roughly the ridge of the H I emission from the 25 arc s map of Fig. 1. Finally, the main features of the H β (dashed contours) and the H I (solid contours) along with the dust lanes (thin and thick lines) are summarized in Fig. 2c.

Figure 2a shows that the H β emission (and thus the young, massive stars which ionize the gas) appears along a locus roughly parallel to the dust lane, but shifted about 15 arc s to the east of it. This shift corresponds to a radial displacement of 700 pc in the plane of the galaxy for an assumed distance of 8.9 Mpc and inclination of 24°.

Figure 2b shows that the H I ridge is also displaced to the east of the dust lane, and the approximate coincidence of the dashed line (the 25 arc s H I ridge) with the peaks of H I in Fig. 2b (10 arc s resolution) confirms that this shift is not a consequence of beam-smoothing.

Fig. 2c shows that, while the ridge line of the H I (solid contour) corresponds closely with that of the H β (dashed contours), there is no detailed coincidence between the individual bright peaks of atomic and ionized hydrogen.

In the framework of the picture (summarized in ref. 12) in which a passing density wave in the underlying stellar disk induces shocks in the gas followed by cloud collapse and star formation, several inferences concerning changes in the state of the interstellar gas in M83 can now be made. In this theoretical picture, the position of the shock front is given by the dust lane, and the typical gas streaming into and out of this region is indicated by the numbered arrows 1 and 2 in Fig. 2a (the observed structure of the velocity field¹³ indicates that the south-

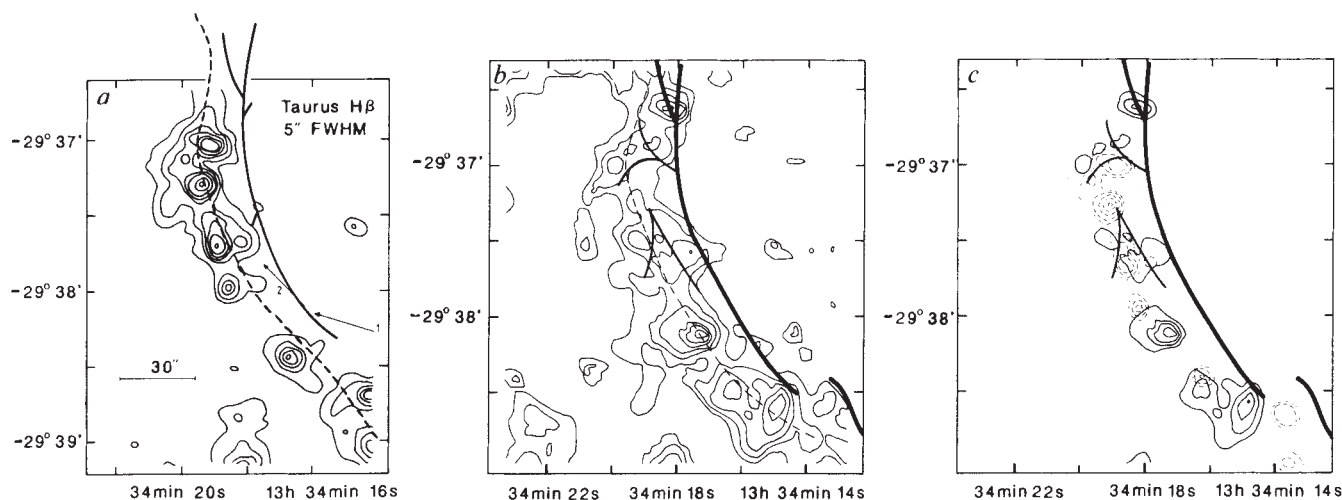


Fig. 2 *a*, The distribution of dust (thick solid line), $H\beta$ emission (contours, 5 arc s resolution), and the ridge line of the $H\text{ I}$ from the contours in Fig. 1 (dashed line) are shown here for a section of the inner eastern spiral arm in M83. The horizontal bar at the lower left indicates 30 arc s, corresponding to 530 pc or 1,280 pc along the major axis of M83 for assumed distances of 3.7 and 8.9 Mpc, respectively. The arrows marked 1 and 2 indicate the pre- and post-shock gas flow. *b*, Contours of surface density of atomic hydrogen at 10 arc s resolution, along with the $H\text{ I}$ ridge line (dashed) from the 25 arc s map of Fig. 1. *c*, Peaks of $H\text{ I}$ (solid contours), $H\beta$ (dashed contours) and dust lanes in the eastern arm of M83.

east side of the galaxy is the far side if the spiral arms are trailing). Because Figure 2*b* shows essentially no atomic hydrogen in the neighbourhood of the dust lane, the interstellar gas there is apparently mostly molecular; furthermore, it was molecular before coming into the shock, and it remains molecular for some time thereafter. The timescales depend on the details of the actual flow velocities and the speed of the underlying spiral pattern; further analysis of the observations will provide more accurate estimates. Figure 2*b* shows the $H\text{ I}$ gas appearing in quantity quite abruptly in the region downstream from the shock where the young stars have generated enough ionizing photons to produce the $H\text{ II}$ complexes visible by the $H\beta$ emission. Note that the $H\text{ I}$ appears also between the $H\text{ II}$ complexes, indicating that young stars have been formed there as well; these stars are apparently hot enough to dissociate the H_2 ($h\nu \geq 11.2\text{ eV}$) but not hot enough to ionize it extensively ($h\nu > 13.6\text{ eV}$). Furthermore, the approximate anti-coincidence of $H\text{ I}$ and $H\beta$ peaks in Fig. 2*c* confirms that the timescale during which the interstellar gas changes state from molecular to atomic to ionized is short compared with the expected time for cloud collapse in spiral arms¹⁷ of $30\text{--}50 \times 10^6\text{ yr}$ and to the observed lifetime of the ionizing stars in young associations¹⁸ of $\sim 10^7\text{ yr}$. The presence of highly-obscured $H\text{ II}$ complexes embedded in the $H\text{ I}$ clouds cannot be excluded by the optical data because they would be invisible in the $H\beta$ observations. Radio continuum observations at several frequencies with sufficient sensitivity and angular resolution would allow identification of such complexes from their thermal emission. A preliminary study of M83 with the VLA by M. Ondrechen (personal communication) at 21 and 6 cm reveals a smooth, nonthermal ridge of radio continuum emission coincident with the dust lane in this region of the galaxy, confirming that the dust lane not only indicates a local maximum in the interstellar gas density but also that it does not hide a substantial thermal contribution to the radio emission. Unfortunately, Ondrechen's observations are not sensitive enough to detect the thermal emission from the less prominent $H\text{ II}$ complexes, although several of the optically brightest objects do appear on his maps. Radio studies of other galaxies^{19,20} confirm that the bright thermal radio sources always have optically visible counterparts in the same area on a scale of 0.5 kpc, although the detailed distributions of radio and optical surface brightness can be

different as a result of nonuniform extinction on finer scales (as is also the case for the much smaller $H\text{ II}$ regions in the Galaxy). We, therefore, consider it unlikely that the $H\beta$ emission has failed to reveal major star-forming regions in M83.

Finally, in this scenario, the gas must eventually become molecular again further downstream. This occurs because in the density wave picture the gas will eventually reappear near streamline 1 in Fig. 2*a* after traversing other spiral arms in its trip around the Galaxy. Depending on the amounts of dust and less massive ultraviolet-producing stars, the timescale for this recombination will probably be longer than for the dissociation-ionization processes.

We emphasize that the region in M83 which we have studied is exceptional in that the displacements described and analysed above are so clear. Our interpretation of this fact is that, for this specific region, the chaotic motions are apparently small enough to allow the well-ordered density-wave streaming velocities to stretch out the time sequence of star formation into a clear spatial separation. Such large-scale ordered motions in spiral arms have been identified previously, for example, in M51 (ref. 15) and M81 (ref. 16). The lack of such a clear separation in other parts of M83 does not contradict our arguments; the chaotic motions apparently dominate over the ordered streaming there. This picture of the kinematics is basic to our interpretation and can be tested by a detailed examination of the motions of the $H\text{ I}$ and $H\text{ II}$ gas in all the spiral arms of M83. Such an analysis is in progress.

The photo-dissociation of molecular gas by ultraviolet radiation from newly formed stars is a viable explanation for features observed in the Galactic interstellar medium on the length scale of one parsec. Our observational results support the suggestion that the same process can also be ordered by spiral structure and substantially affect the $H\text{ I}$ morphology of galaxies on the kpc scale. Furthermore, our results indicate that $H\text{ I}$ clouds need not always have a key role as precursors of molecular clouds, as is thought to be the case in the current picture of the star formation sequence. The relationship between atomic and molecular gas in galaxies may be more subtle and significantly affected by the star formation process itself.

We thank R. Humphreys for an optical photograph of M83, M. Ondrechen and J. M. van der Hulst for permission to use their VLA- $H\text{ I}$ data, and W. M. Goss for discussions.

Received 10 July; accepted 18 October 1985.

- Spitzer, L. *Physical Processes in the Interstellar Medium*, 122-127 (Wiley, New York, 1978).
- Hollenbach, D. J., Werner, M. W. & Salpeter, E. E. *Astrophys. J.* **163**, 165-180 (1971).
- Hill, J. K. & Hollenbach, D. J. *Astrophys. J.* **225**, 390-404 (1978).
- Read, P. L. *Mon. Not. R. astr. Soc.* **192**, 11-32 (1980).
- Read, P. L. *Mon. Not. R. astr. Soc.* **194**, 863-878 (1981).
- Goss, W. M., Retallack, D. S., Felli, M. & Shaver, P. A. *Astr. Astrophys.* **117**, 115-126 (1983).
- Young, J. S. & Scoville, N. Z. *Astrophys. J.* **258**, 467-489 (1982).
- Scoville, N. Z. & Young, J. S. *Astrophys. J.* **265**, 148-165 (1983).
- Combes, F., Encrenaz, P. J., Lucas, R. & Welachew, L. *Astr. Astrophys.* **67**, L13-L15 (1978).
- Rickard, L. J. & Blitz, L. *Astrophys. J. Lett.* **292**, L57-L60 (1985).
- Atherton, P. D. *et al. Mon. Not. R. astr. Soc.* **201**, 661-696 (1982).
- Shu, F. H. *The Physical Universe*, 281-284 (University Science Books, Mill Valley, California, 1982).
- Comte, G. *Astr. Astrophys. Suppl.* **44**, 441-450 (1981).
- Rogstad, D. H., Lockhart, I. A. & Wright, M. C. H. *Astrophys. J.* **193**, 309-319 (1974).
- Segalovitz, A. *Astr. Astrophys.* **52**, 167-174 (1976).
- Visser, H. *Astr. Astrophys.* **88**, 159-174 (1980).
- Elmegreen, B. G. *Astrophys. J.* **231**, 372-383 (1979).
- Blaauw, A. A. *Rev. Astr. Astrophys.* **2**, 213-246 (1962).
- Israel, F. P., Goss, W. M. & Allen, R. J. *Astr. Astrophys.* **40**, 421-438 (1975).
- Van der Kruit, P. C., Allen, R. J. & Rots, A. H. *Astr. Astrophys.* **55**, 421-433 (1977).
- Field, G. B. *Proc. Inst. Radio Engin. Aust.* **46**, 240-250 (1958).

Astrophysical implications of amorphous ice—a microporous solid

Erwin Mayer & Rudolf Pletzer

Institut für Anorganische und Analytische Chemie,
Universität Innsbruck, A-6020 Innsbruck, Austria

Since the observation of the 3- μm band in interstellar infrared sources¹⁻³, vapour-deposited amorphous ice, $\text{H}_2\text{O}(\text{as})$, has been discussed as a major component of comets⁴⁻⁹, of satellites of the outer planets^{10,11} and of interstellar dust¹¹⁻¹³. Some of the physical properties of $\text{H}_2\text{O}(\text{as})$ important for these discussions appear to be contradictory: for example N_2 adsorption isotherms at 77 K evaluated by the BET method¹⁴ gave surface areas of $241 \text{ m}^2 \text{ g}^{-1}$ according to Ghormley^{14,15}, but less than $12 \text{ m}^2 \text{ g}^{-1}$ according to others^{16,17}. We have reinvestigated adsorption of N_2 on $\text{H}_2\text{O}(\text{as})$ and find that it is a microporous solid. The differences between adsorption phenomena taking place in micropores and on the surface of intermediate pores or non-porous adsorbents should be of general importance for the various speculations of the role of $\text{H}_2\text{O}(\text{as})$ in space. In particular, we discuss the influence of micropores on the H_2 recombination rate on $\text{H}_2\text{O}(\text{as})$ in interstellar dust and on adsorption of volatile gases in comets.

Micropores are defined as pores of $<20 \text{ \AA}$ (ref. 18) or $<18 \text{ \AA}$ (ref. 19) width. A characteristic feature²⁰ of adsorption on microporous adsorbents is the substantial increase in the adsorption energy, and consequently, in the adsorption potentials in micropores compared with the corresponding values for large-porous, intermediate-porous or non-porous adsorbents. Model calculations further showed that adsorption in pores of molecular diameters is sufficiently different from that in coarser pores to justify their assignment to a separate category as micropores²¹.

In previous N_2 isotherm investigations, $\text{H}_2\text{O}(\text{as})$ has been prepared with fairly high deposition rates¹⁴⁻¹⁷. These conditions lead to supersonic flow of the water vapour, with the exit of a capillary acting as a nozzle. We have recently shown that supersonic flow has a strong influence on some of the physical properties of $\text{H}_2\text{O}(\text{as})$ (refs 22, 23). To exclude such effects and to imitate more closely the conditions in space, we chose the lowest deposition rate experimentally feasible, this giving molecular flow of the water vapour. Figure 1A shows a N_2 adsorption isotherm on $\text{H}_2\text{O}(\text{as})$ which was prepared by admitting water vapour through a nozzle into a high-vacuum system and depositing the vapour on the flat-bottom part of a cool finger, thus acting as a cryoplate. The apparent surface area of isotherm A was calculated by BET to be $421 \text{ m}^2 \text{ g}^{-1}$. In a second experiment (Fig. 1D) a lower apparent surface area of $352 \text{ m}^2 \text{ g}^{-1}$ was obtained for the same rate and time of water vapour deposi-

tion. In this experiment, the distance between nozzle exit and cryoplate had been increased from 12 (for isotherm A) to 82 mm, and some $\text{H}_2\text{O}(\text{as})$ had condensed in the upper, warmer part of the cool finger, thus explaining the lower value. Microporous solids having relatively small external surfaces are characterized by a type I isotherm which is concave to the p/p^0 axis with a horizontal plateau^{19,20}. However, more often, microporosity is associated with an appreciable external surface, and a composite isotherm of type II is observed containing contributions from micropores (type I) and the external surface (type II)¹⁹; isotherms A and D in Fig. 1 are typical examples of this.

We have determined the micropore volume component in these isotherms following the recent IUPAC recommendations²⁴: the procedure for assessing micropore volume is either graphical comparison of an experimental isotherm with a non-porous standard isotherm, or evaluation by Dubinin-Radushkevich (DR) plot²⁵. The appropriate standard isotherm for evaluation of microporosity is still being debated. We followed Sing²⁶ and the IUPAC recommendations²⁴ in choosing as a reference an isotherm obtained on a chemically similar solid. More specifically, the standard isotherm selected was that obtained on $\text{H}_2\text{O}(\text{as})$ deposited at 113 K, the highest deposition temperature at which simultaneous formation of ice I is still small (Fig. 1B, apparent surface area, $40 \text{ m}^2 \text{ g}^{-1}$). We do not know whether $\text{H}_2\text{O}(\text{as})$ formed by deposition at 113 K is completely non-porous, but we assume that the BET surface area of $40 \text{ m}^2 \text{ g}^{-1}$ is totally external; additional pores in this reference solid would give a lower value for the micropore volume. Subtraction of the standard isotherm B from curve A in Fig. 1 gives curve C, which is typical of a type I isotherm of a purely microporous solid¹⁹.

The graphical evaluation of the micropore volumes is shown in Fig. 2. The comparison plots in Fig. 2a are simple plots of uptake per unit mass of the sample materials (in mmol g^{-1}) against that of the reference material at the same relative pressure^{27,28}. The micropore volumes are determined from the extrapolated ordinate intercepts, which gives micropore volumes of $0.20 \text{ cm}^3 \text{ g}^{-1}$ for curve A and $0.17 \text{ cm}^3 \text{ g}^{-1}$ for curve D (using 0.808 g cm^{-3} as the density of liquid N_2 at 77 K). From the slopes external surface areas of $43 \text{ m}^2 \text{ g}^{-1}$ and $84 \text{ m}^2 \text{ g}^{-1}$ are obtained for A and D, respectively. Evaluation by DR plots in Fig. 2b, where the log of the amount adsorbed is plotted against $\log^2(p^0/p)$ where p^0 is the saturation pressure of the pure adsorbate at the temperature of the measurement, gave micropore volumes of $0.21 \text{ cm}^3 \text{ g}^{-1}$ for A and $0.12 \text{ cm}^3 \text{ g}^{-1}$ for D. For isotherm A, the agreement between the micropore volumes obtained by two different methods is better than usual and probably coincidental; for isotherm D it is more typical, the DR plot values often being much lower²⁹. A more detailed isotherm analysis is not meaningful because, at present, there is no reliable procedure for the computation of micropore size distribution from a single isotherm²⁴. However, according to Gregg and Sing¹⁹, the gradual approach of Fig. 1C to the plateau region is an indication of many larger micropores (7-18 $\text{\AA}$ width) which are filled by a mainly cooperative process.

We have found no evidence for hysteresis loops during desorption. This indicates that in our $\text{H}_2\text{O}(\text{as})$ samples the fraction of mesopores, defined as being between 20 and 500 $\text{\AA}$ width^{19,24}, must have been only small. This is supported by the comparison plots in Fig. 2a because the presence of mesopores in a microporous solid would have been recognizable by an upward deviation from linearity at $p/p^0 > 0.4$ (ref. 19).

A large micropore volume agrees fairly well with the density value of $0.81 \pm 0.02 \text{ g cm}^{-3}$ obtained by Seiber *et al.*³⁰ *in situ* using optical thin-film interference techniques. However, this seems to be inconsistent with the reported density value of $\text{H}_2\text{O}(\text{as})$ (ref. 31) of $0.94 \pm 0.02 \text{ g cm}^{-3}$, as pointed out by Sceats and Rice^{32,33}. From X-ray data the density of $\text{H}_2\text{O}(\text{as})$ was estimated to be $0.94 \pm 0.02 \text{ g cm}^{-3}$ (ref. 33). However, as pointed out by Temkin *et al.*³⁴ density variations may arise from a change

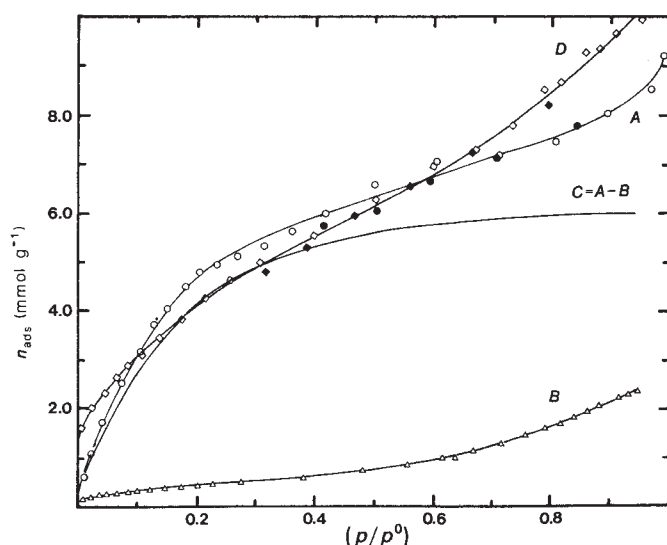


Fig. 1 N_2 -adsorption isotherms (static volumetric at 77 K) on (A) and (D), $H_2O(as)$ prepared by vapour deposition at 77 K and at 113 K (B); C, isotherm obtained by subtracting B from A. Open symbols, adsorption; solid symbols, desorption. (BET values (with $a_m(N_2) = 16 \text{ \AA}^2$ where a_m is the average area occupied by an adsorbed N_2 molecule in the complete monolayer): for A, apparent surface area, $a_s = 421 \text{ m}^2 \text{ g}^{-1}$, $c = 18$; for B, $a_s = 40 \text{ m}^2 \text{ g}^{-1}$, $c = 26$; for D, $a_s = 352 \text{ m}^2 \text{ g}^{-1}$, $c = 30$, where a_s is the apparent surface area and c the BET constant). Experimental details: $H_2O(as)$ was prepared in a sort of sublimation apparatus made of glass, H_2O vapour entering through a fine metering valve and nozzle with 4 mm diameter at the bottom of the apparatus. For $H_2O(as)$ formation only the cool finger was cooled, the flat bottom part of the finger with 10 cm^2 area acting as cryoplate. For isotherm measurements, the whole apparatus was immersed in liquid N_2 and adsorption and desorption measured *in situ*. It is important to add gaseous N_2 (99.9995 %) very slowly at each adsorption step to avoid local heating and annealing of the sample. Equilibration for A and D needed 5–10 min, but for B was nearly instantaneous. A large volume of the apparatus (457 cm^3) was necessary for a good vacuum and unobstructed gas flow during $H_2O(as)$ formation, allowing only surface areas $> 10 \text{ m}^2$ to be measured. Apparatus and measurement techniques were tested with a sample of active carbon of known surface area. Details of $H_2O(as)$ preparation: A, $p^0 (H_2O)$ before nozzle $= 9 \times 10^{-4} \text{ mbar}$, $K_n = 11$ where K_n , the Knudsen number, is the ratio of mean free path in the gas supply to the smallest dimension of the orifice 137 h deposition time, base pressure during deposition $< 10^{-6} \text{ mbar}$, 0.133 g sample of $H_2O(as)$; a short distance of 12 mm between nozzle exit and cryoplate was selected to avoid deposition of $H_2O(as)$ on the upper, warmer part of the cool finger; from the area of the flat bottom part of the cool finger (10 cm^2) and the weight (g) of $H_2O(as)$, an average thickness of 0.15 mm is obtained. This gives an approximate deposition rate of $3 \text{ \AA s}^{-1}$. B, 113 K cryoplate temperature with isopentane N_2 slush, $p^0 = 0.04 \text{ mbar}$, $K_n = 0.080$, 6 h deposition time, 1.70 g $H_2O(as)$. X-ray diffraction showed that 113 K deposition gives largely $H_2O(as)$, with a small amount of ice I only. D, Identical to A except for the longer distance of 82 mm between nozzle and cryoplate, 143 h deposition time and 0.153 g formed $H_2O(as)$.

in the number and distribution of voids, a change in the atomic network, or a combination of the two. Moreover, it needs measurements at smaller angles than used by Narten *et al.*³³, where scattering from voids predominates, to determine number and distribution of voids or pores. We have redetermined the density of a microporous sample of $H_2O(as)$, using a flotation method involving hydrometer in a liquid N_2/O_2 mixture, and obtained a value of $0.910 \pm 0.005 \text{ g cm}^{-3}$. This value would still be too high for the micropore volumes. This apparent inconsistency stems from the fact²³ that in $H_2O(as)$ samples of high apparent surface area, the volume is reduced to $\sim 50 \text{ m}^3 \text{ g}^{-1}$ when such samples are heated for several minutes to 113 K, or possibly below that temperature. As it is impossible to avoid

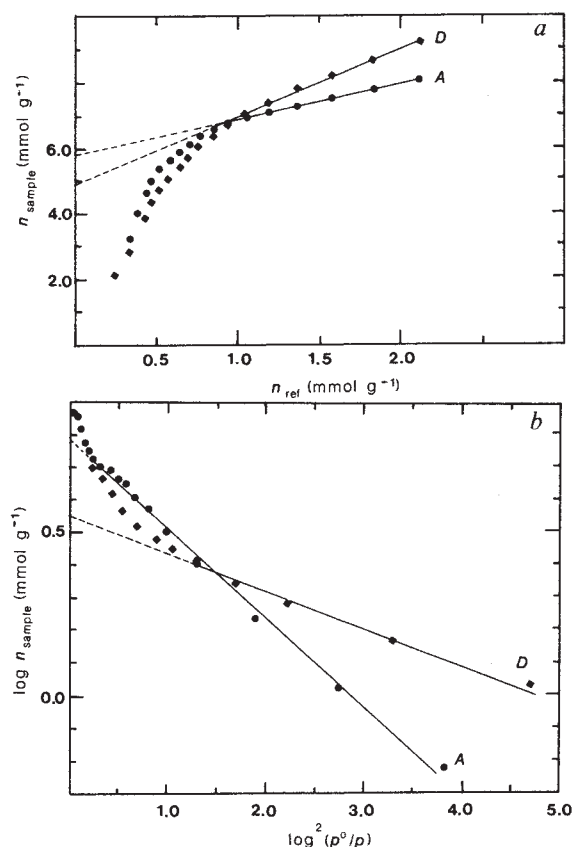


Fig. 2 Evaluation of micropore volume in $H_2O(as)$ used for curves A and D in Fig. 1. a, $0.20 \text{ cm}^3 \text{ g}^{-1}$ (for A) and $0.17 \text{ cm}^3 \text{ g}^{-1}$ (for D) obtained by comparison plots, from a plot of mmol adsorbed N_2 per g $H_2O(as)$ of sample against reference B at the same p/p^0 values; external surfaces of $43 \text{ m}^2 \text{ g}^{-1}$ (for A) and $84 \text{ m}^2 \text{ g}^{-1}$ (for D) were calculated from the slopes. b, $0.21 \text{ cm}^3 \text{ g}^{-1}$ (for A) and $0.12 \text{ cm}^3 \text{ g}^{-1}$ (for D) by DR plots.

some warming of $H_2O(as)$ during the various manipulations for density measurements by flotation, annealed samples with comparatively low surface areas must have been investigated. This anomaly is supported by our observation that when, in the hope of obtaining higher precision with a smaller volume, $H_2O(as)$ in liquid N_2 is transferred very carefully into a small tube for adsorption measurements, annealed, although according to X-ray diffraction still amorphous, samples are obtained which have surface areas of $15\text{--}60 \text{ m}^2 \text{ g}^{-1}$ instead of $200\text{--}400 \text{ m}^2 \text{ g}^{-1}$. Therefore, *in situ* measurement was essential for investigating surface properties of non-annealed $H_2O(as)$. The same explanation probably holds for the low surface area values of $H_2O(as)$ reported^{16,17}. Comparable annealing behaviour and density variations have been reported for vapour-deposited amorphous Ge and Si (refs 35, 36), both of which are believed to be structurally related to $H_2O(as)$ (refs 32, 37), and have been interpreted in terms of a network of small voids coalescing at higher temperature³⁵.

We can only speculate on the shape of the isotherm of a sample of $H_2O(as)$ prepared and investigated at lower temperatures. Ghormley¹⁴ had predicted that $H_2O(as)$ as prepared at 20 K should have a surface area about double his value of $241 \text{ m}^2 \text{ g}^{-1}$ at 77 K. We had recently obtained independent evidence for irreversible annealing effects during warm-up of $H_2O(as)$ from 16 to 77 K from halving of the bandwidth of the decoupled O–D stretching transition³⁸. Although the band-width of the decoupled oscillator as an intrinsic property is not directly related to surface area or micropore volume, it certainly indicates considerable annealing effects at extremely low temperatures. Therefore, by extrapolation we expect that deposition tem-

peratures below 77 K should produce even larger apparent surface areas and micropore volumes.

The evidence for large numbers of micropores in $\text{H}_2\text{O}(\text{as})$ prepared with very low rates of deposition should be of general importance for physical adsorption on $\text{H}_2\text{O}(\text{as})$ in space, on the premises that it really exists, because both the equilibrium and the kinetics of adsorption are strongly influenced by the presence of micropores^{19,20,39}. We will discuss two areas where consideration of micropores should influence the existing concepts of the role of $\text{H}_2\text{O}(\text{as})$ in space.

First, the rate of formation of H_2 molecules from adsorbed H atoms on interstellar ice grains, an efficient process on single crystal ice according to Hollenbach and Salpeter^{40,41}, could be orders of magnitude lower on $\text{H}_2\text{O}(\text{as})$ according to Smoluchowski¹¹⁻¹³, mainly because in his analysis only H atoms accidentally adsorbed on neighbouring sites can form H_2 molecules. This might be different for microporous $\text{H}_2\text{O}(\text{as})$ where the primary effect is volume filling of the micropores already at very low relative pressures, the smallest pores being filled first⁴². Relative to $\text{H}_2\text{O}(\text{as})$ lacking micropores, for a given temperature and gas pressure the residence time of the adsorbed species is strongly increased. Qualitatively, one might expect that in the micropores there is an increased probability of an encounter between two H atoms and the formation of H_2 , whether the adsorbed species are immobilized in the pores or are still able to migrate. A micropore effect as discussed above obviously presumes pure physical adsorption for atomic hydrogen on $\text{H}_2\text{O}(\text{as})$. A quantitative description, assuming micropores of various shapes and sizes in line with Everett and Powl²¹, should be possible. Experimental data on the sticking probability of H atoms⁴³ and H_2 (ref. 44) and on recombination of atomic hydrogen⁴⁵ on icy surfaces have been reported. Because these surfaces were not well defined, an investigation of the recombination on microporous $\text{H}_2\text{O}(\text{as})$ prepared as described here would be of interest.

Second, microporous $\text{H}_2\text{O}(\text{as})$ might be of importance in cometary nuclei, and a combination of its adsorption and annealing properties could explain some of the surprising properties of comets. The enormous gas production observed in comets led to the development by Whipple⁴⁶, of the 'icy conglomerate model', where the main constituent of the nucleus is believed to be water ice. To account for the fact that the fuzzy head of comets, although containing gases of very different volatility, typically appears at distances of 3 AU, Delsemme and Swings⁴⁷ first suggested vaporization of clathrate hydrates. Miller⁴⁸ had pointed out that mixed clathrate hydrates are required to account for the constant ratios of gaseous species because the gases from a single hydrate crystal probably all evaporate at the same time and not as some function of their binding energies in the cavity. The problem with the clathrate hydrate model is that thermodynamic stability alone is not sufficient; the kinetics of formation must also be considered: CH_4 (ref. 49) and Ar (ref. 50) clathrate hydrates were reported to form at 82 and 90 K, respectively, while all other clathrate hydrates need much higher temperatures. It is at least doubtful whether the rate of clathrate hydrate formation is high enough at very low temperatures, even considering astronomical time-scales. If microporous $\text{H}_2\text{O}(\text{as})$ is considered to be a component of cometary nuclei, there is no problem about storing large amounts of gas by adsorption. A further peculiar property of microporous $\text{H}_2\text{O}(\text{as})$ is that during warming of $\text{H}_2\text{O}(\text{as})$ in the presence of adsorbed gas to 113 K, large amounts of gas are enclosed which cannot be pumped off at temperatures ≤ 113 K and 10^{-6} mbar pressure: the amount of enclosed gas in mmol g^{-1} $\text{H}_2\text{O}(\text{as})$ was 0.62 for N_2 and 2.73 for O_2 , corresponding to respective $\text{H}_2\text{O}/\text{gas}$ ratios of 90 and 20. We have previously reported the reduction of the surface area of $\text{H}_2\text{O}(\text{as})$ to $\sim 50 \text{ m}^2 \text{ g}^{-1}$; in the presence of adsorbed gas, simultaneous enclosure of gas in the filled micropores occurs. We believe that the amount of enclosed gas is determined by two competing

effects, desorption of gas during warm-up and reduction of surface area by pore closure in the same temperature regime. This explains why the amount of enclosed O_2 as the less volatile gas is higher by a factor of ~ 5 , the $\text{H}_2\text{O}/\text{gas}$ ratio already approaching the ratio in structure II clathrate hydrates; this represents a fractionation, the most volatile components being lost first, as postulated for comets⁹. The X-ray diffractogram of $\text{H}_2\text{O}(\text{as})$ before and after gas enclosure was identical, thus proving that no devitrification or formation of crystalline clathrate hydrate had occurred. The $\text{H}_2\text{O}(\text{as})/\text{enclosed gas}$ samples might resemble what have sometimes been called 'amorphous clathrate ices'. These experiments differ from those of Ghormley¹⁴ in that he had condensed H_2O and O_2 vapour simultaneously to $\text{H}_2\text{O}(\text{as})$ with enclosed O_2 ; however, they are similar in that vaporization of the enclosed gas occurred slowly over a wide temperature range. To obtain sufficient material, we had to prepare $\text{H}_2\text{O}(\text{as})$ for the gas enclosure experiments using much higher rates of deposition than for the adsorption isotherms. However, the $\text{H}_2\text{O}(\text{as})$ samples seem to be very similar with respect to micropore volume and annealing behaviour.

Applied to the cometary nucleus, the microporous $\text{H}_2\text{O}(\text{as})$ could store large amounts of gases of very different volatility which might have a similar behaviour during their approach to the Sun as a crystalline mixed clathrate hydrate, giving off gases of very different volatilities with fairly constant ratios. The advantage of this model compared with the postulated formation of crystalline clathrate hydrates is that large amounts of volatile gas can be enclosed in the micropores even at very low temperatures, possibly even far below 77 K, our lowest experimental temperature, either by thermal annealing and closure of filled micropores or by pore closure due to further layers of deposited $\text{H}_2\text{O}(\text{as})$. Depending on the temperature and the rate of nucleation or crystalline clathrate hydrate formation, either the enclosed volatile gases could vaporize directly, or devitrification of $\text{H}_2\text{O}(\text{as})$ to cubic ice and/or crystalline clathrate hydrate formation could occur. Although these are only speculations, many of the experimental details implied can be investigated and tested in laboratory experiments. Perhaps the most drastic simplification, not amenable to experiment, is the extrapolation from microporous $\text{H}_2\text{O}(\text{as})$ samples of, at most, 2 g to the sphere of a cometary nucleus of ~ 1 km diameter, because annealing behaviour and surface properties should be quite different in a 'dirty snowball' of this size.

This work was supported by the Fonds zur Förderung der wissenschaftlichen Forschung of Austria. We thank Dr R. Kramer and Professor E. Bechtold for discussions.

Received 17 June; accepted 17 October 1985.

1. Gillett, F. C. & Forrest, W. J. *Astrophys. J.* **179**, 483 (1973).
2. Gillett, F. C., Jones, T. W., Merrill, K. M. & Stein, W. A. *Astr. Astrophys.* **45**, 77-81 (1975).
3. Merrill, K. M., Russell, R. W. & Soifer, B. T. *Astrophys. J.* **207**, 763-769 (1976).
4. Patashnick, H., Rupprecht, G. & Schuerman, D. W. *Nature* **250**, 313-314 (1974).
5. Klinger, J. *Science* **209**, 271-272 (1980).
6. Smoluchowski, R. *Astrophys. J. Lett.* **244**, L31-L34 (1981).
7. Klinger, J. *Icarus* **47**, 320-324 (1981).
8. Klinger, J. *J. phys. Chem.* **87**, 4209-4214 (1983).
9. Delsemme, A. H. *J. Phys. Chem.* **87**, 4214-4218 (1983).
10. Smoluchowski, R. *Science* **201**, 809-811 (1978).
11. Smoluchowski, R. *Astrophys. Space Sci.* **65**, 29-38 (1979).
12. Smoluchowski, R. *Astrophys. Space Sci.* **75**, 353-363 (1981).
13. Smoluchowski, R. *J. phys. Chem.* **87**, 4229-4233 (1983).
14. Ghormley, J. A. *J. chem. Phys.* **46**, 1321-1325 (1967).
15. Ghormley, J. A. *J. chem. Phys.* **48**, 503-508 (1968).
16. Adamson, A. W., Dormant, L. M. & Oram, M. J. *J. Colloid Interface Sci.* **25**, 206-217 (1967).
17. Ocampo, J. & Klinger, J. *J. Colloid Interface Sci.* **86**, 377-383 (1982).
18. Everett, D. H. *Pure appl. Chem.* **31**, 579-638 (1972).
19. Gregg, S. J. & Sing, K. S. W. *Adsorption, Surface Area and Porosity* (Academic, New York, 1982).
20. Dubinin, M. M. *J. Colloid Interface Sci.* **23**, 487-499 (1967); *Prog. Surface membrane Sci.* **9**, 1-70 (1975).
21. Everett, D. H. & Powl, J. C. *JCS Faraday Trans. I* **72**, 619-636 (1976).
22. Mayer, E. & Pletzer, R. *J. chem. Phys.* **80**, 2939-2952 (1984).
23. Mayer, E. & Pletzer, R. in *Ices in the Solar System* (eds Klinger, J., Benest, D., Dollfus, A. & Smoluchowski, R.) 81-88 (Reidel, Dordrecht, 1985).
24. Sing, K. S. W. *Pure appl. Chem.* **54**, 2201-2218 (1982).
25. Dubinin, M. M. & Radushkevich, L. V. *Proc. natn. Acad. Sci. USSR* **55**, 331 (1947).
26. Sing, K. S. W. *Chem. Ind.* 829-830 (1967).
27. Brown, L. E. & Hall, P. G. *Trans. Faraday Soc.* **67**, 3558-3564 (1971).
28. Lee, J. A. & Newnham, C. E. *J. Colloid Interface Sci.* **56**, 391-394 (1976).

29. Marsh, H. & Rand, B. in *3rd Conf. Industrial Carbon and Graphite*, 93 (Academic, London, 1970).
30. Seiber, B. A., Wood, B. E. & Smith, A. M. *Science* **170**, 652-654 (1970).
31. Ghormley, J. A. & Hochanadel, C. J. *Science* **171**, 62-64 (1971).
32. Seatz, M. G. & Rice, S. A. in *Water, a Comprehensive Treatise* Vol. 7 (ed. Franks, F.) Ch. 2 (Plenum, New York, 1982).
33. Narten, A. H., Venkatesh, C. G. & Rice, S. A. *J. chem. Phys.* **64**, 1106-1121 (1976).
34. Temkin, R. J., Paul, W. & Connell, G. A. *Adv. Phys.* **22**, 581-641 (1973).
35. Ehrenreich, H. & Turnbull, D. *J. Non-Crystalline Solids* **8-10**, 19-35 (1972).
36. Bienenstock, A. *Amorphous Liquid Semiconductors, Proc. 5th int. Conf.* (Taylor & Francis, London, 1974).
37. Alben, R. & Boutron, P. *Science* **187**, 430-432 (1975).
38. Mayer, E. & Pletzer, R. *J. chem. Phys.* (in the press).
39. Aharoni, C. & Suzin, Y. in *Characterization of Porous Solids* (eds Gregg, S. J., Sing, K. S. W. & Stoeckli, H. F.) 173-180 (Society of Chemical Industry, London, 1979).
40. Hollenbach, D. & Salpeter, E. E. *J. chem. Phys.* **53**, 79-86 (1970).
41. Hollenbach, D. & Salpeter, E. E. *Astrophys. J.* **163**, 155-164 (1971).
42. Poncet, V., Knor, Z. & Černý, S. *Adsorption on Solids*, 373 (Butterworths, London, 1974).
43. Brackmann, R. T. & Fite, W. L. *J. chem. Phys.* **34**, 1572-1579 (1961).
44. Govers, T. R., Mattera, L. & Scoles, G. *J. chem. Phys.* **72**, 5446-5455 (1980).
45. Schutte, A., Bassi, D., Tommasini, F. & Turelli, A. *J. chem. Phys.* **64**, 4135-4142 (1976).
46. Whipple, F. L. *Astrophys. J.* **111**, 375-394 (1950), **113**, 464-474 (1951).
47. Delsemme, A. H. & Swings, P. *Ann. Astrophys.* **15**, 1-6 (1952).
48. Miller, S. L. *Proc. natn. Acad. Sci. U.S.A.* **47**, 1798-1808 (1961).
49. Delsemme, A. H. & Wenger, A. *Planet. Space Sci.* **18**, 709-715 (1970).
50. Barrer, R. M. & Edge, A. V. *Proc. R. Soc. A* **300**, 1-24 (1967).

Importance of local mesoscale factors in any assessment of nuclear winter

B. W. Golding, P. Goldsmith*, N. A. Machin & A. Slings

Meteorological Office, London Road, Bracknell, Berkshire RG12 2SZ, UK

There is a possibility that severe climate perturbations would follow a major nuclear war (the 'nuclear winter') due to the injection of large amounts of smoke into the atmosphere¹⁻⁶. Given assumptions about the amount of smoke generated, its carbon content, the size of the particles and their optical properties, calculations of the subsequent cooling of the surface air temperatures have been made with a variety of atmospheric models. It has not been possible in these models, however, to represent any effects due to the initially patchy nature of the smoke. We investigate here whether such effects could be important by using a 15-km resolution atmospheric model to simulate the mesoscale response to a single smoke plume. Our results show that vertical circulations created by the radiative heating of the smoke may be sufficiently intense to lead to cloud formation and hence to scavenging of the smoke.

Simulations of the cooling of the surface following a nuclear war have been based on either one-dimensional models¹ or versions of coarse three-dimensional global atmospheric circulation models originally designed to study global climate variability. The one-dimensional models have to assume a uniform smoke distribution while the three-dimensional models⁶ have a resolution of no better than about 5° latitude which prevents the representation of mesoscale processes. Given the local nature of the initial smoke intrusions, and the strong temperature gradients that their radiative characteristics would create, it is reasonable to suppose that mesoscale atmospheric circulations would be set up, especially near the edges of the plumes. This is in addition to any circulations that might be expected locally at the source of the smoke due to the heat and moisture which inevitably must be associated with the smoke generation. If such circulations were to produce cloud, it would be expected to increase either the scavenging rate or the rate at which smoke particles coagulate. For a given mass of smoke the radiative effects change significantly as the size of the particles increases⁷.

To test the possibility that such mesoscale circulations might occur, a case study has been completed using a numerical mesoscale model currently under development in the Meteorological Office for improving short-period weather forecasting.

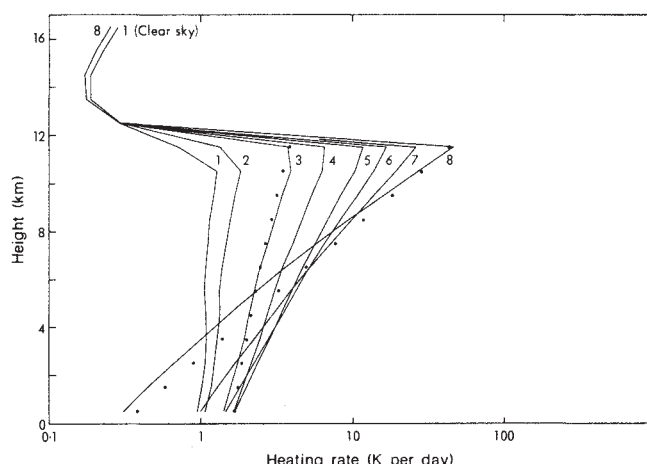


Fig. 1 Vertical profiles of the short-wave heating in an atmosphere containing various loadings of soot aerosols (solid lines). Numbers refer to values given in Table 1. ●, The heating rate profiles obtained with the simple parameterization indicated in Table 1 for curves 3 and 8.

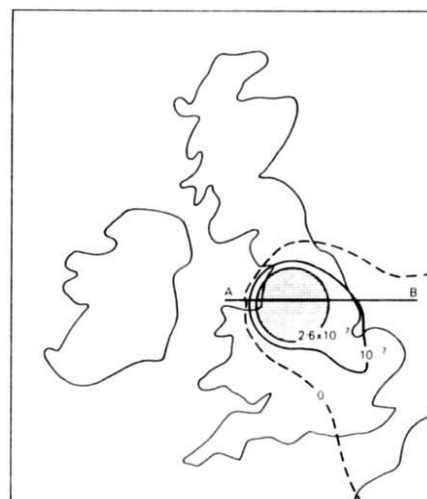


Fig. 2 Smoke concentration in kg per kg at 9 km altitude (~300 mbar) after 12 h integration. The stippled area shows the source region.

Briefly, it is a non-hydrostatic, compressible model using a semi-implicit time-stepping scheme to ensure numerical stability⁸. The vertical coordinate is height above orography⁹, and in the study, 16 levels were used with 5 in the lowest 1 km of the atmosphere and the highest at 12 km. The domain used covered the British Isles with a 15-km grid mesh. The model has an explicit representation of grid-scale clouds¹⁰ and uses a one-dimensional steady-state cloud model to parameterize deep convection. A $1\frac{1}{2}$ -order closure diffusion scheme models sub-grid-scale turbulence. In normal use, radiative effects are calculated only for computing the surface heat balance. A detailed radiation scheme¹¹ was therefore used to develop a parameterization of short-wave heating by smoke which could easily be added to the model. The scheme has 24 spectral bands from 0.2 to 4 μm and treats Rayleigh and multiple scattering, and absorption by smoke, water vapour and ozone. The smoke properties were taken from the baseline calculations in the US National Academy of Sciences⁵. Vertical profiles of the heating rates for various aerosol loadings in a cloudless mid-latitude summer atmosphere, a surface albedo of 0.2 and 60° solar zenith angle

* Present address: European Space Agency, Paris, France.

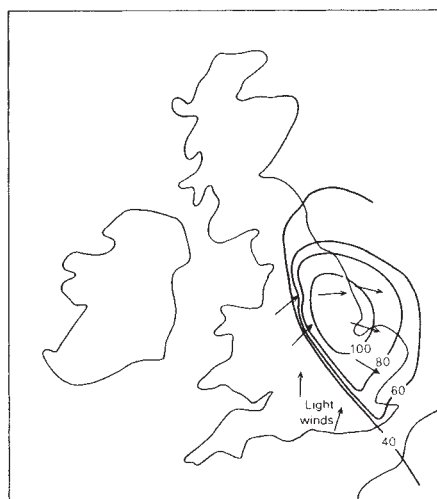


Fig. 3 Air flow and relative humidity (%) at 10.5 km altitude near the smoke plume after 12 h integration.

are plotted in Fig. 1. Downward fluxes at the surface are listed 'exact' in Table 1. In all cases, the heating profile is well approximated by exponential attenuation of the incident flux. It was found that an effective attenuation coefficient of $1.5 \text{ m}^2 \text{ g}^{-1}$ gave a reasonable fit both to the surface flux (last entry in Table 1) and to the heating rates (filled dots) and this was used in the model integration.

In the study, a smoke source was inserted in a column of radius 75 km near the centre of the model grid (see Fig. 2). The vertical profile of the smoke was chosen to be consistent with the published data and to have a concentration of $6.1 \times 10^{-4} \text{ kg m}^{-2}$ over the model depth and a peak at 9 km of $1.1 \times 10^{-7} \text{ kg m}^{-3}$. This profile was artificially constrained to be constant in the source area. The smoke was transported from this source by the model winds but was not affected by any other model variables. The model evolution was affected by the internal radiative heating of the smoke and by the reduction in solar radiation at the ground. Using the absorption coefficient quoted above, we obtained an optical depth of 0.91 for the source region at normal incidence or transmission of 20% of the incoming flux at 53° elevation. This depletion of the solar beam is modest compared with previous studies.

The day chosen for the study was 15 June 1984, when an anticyclone covered England and Wales and skies were generally clear. Initial actual data for 0600 GMT were used. A weak northwesterly wind at upper levels resulted in the main plume spreading to cover much of northern England, the East Midlands

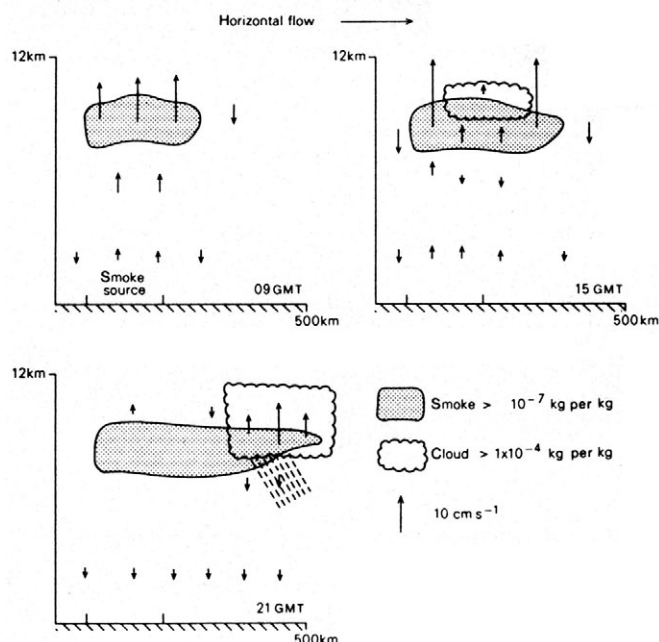


Fig. 4 Schematic representation of the development of a vertical circulation about the smoke and the subsequent formation of cloud seen in the cross-section A-B in Fig. 2.

and East Anglia (see Fig. 2) by 1800 GMT. Solar absorption early in the forecast period quickly resulted in a 5°C temperature excess at about 300 mbar near the centre of the plume compared with a control run for the same occasion but without the smoke plume. This excess produced a vertical circulation of air with maximum upward motion of 0.2 m s^{-1} in the smoke and weaker subsidence around it. One effect of this circulation was to limit the temperature rise through adiabatic cooling, and after sunset to remove it. Another was to increase the humidity as illustrated in Fig. 3, where the air blowing through the source region has been lifted, with the result that its associated water vapour condensed to produce cloud. This cloud was absent in the control run and has a water content sufficient to produce precipitation which, in turn, would evaporate in the dry air below cloud base. The relationship of this cloud to the smoke is shown in the cross-section (Fig. 4) along the line A-B of Fig. 2. Associated with the thickest smoke, upward motion has been much enhanced and, in the indicated regions, cloud has formed. The model does not attempt to represent the physics of the interaction between water and smoke; however, the results confirm that local circulations, capable of producing cloud formations that otherwise would not occur, can be set up by the local smoke plume.

These results suggest that smoke plumes associated with individual targets could have fairly immediate consequences on the local meteorology, with the result that the characteristics of the smoke may be modified before it has been diffused into the continental-scale smoke pall which has been the starting point of 'nuclear winter' calculations in the past. Further work is needed to establish how sensitive this result is to the approximations made in the modelling process, to the weather situation selected, and to the mode of insertion of the smoke, including the interaction of several plumes. The realism of the experiment could be improved, in particular, by a more detailed treatment of the source region, including the concomitant sources of heat and moisture, and also by including the physics of interactions between water and smoke.

In conclusion we emphasize that, in the case described, fairly dry anticyclonic conditions prevailed and so washout of smoke

Table 1 Details of vertical profile curves in Fig. 1

Curve in Fig. 1	Column density (g m^{-2})	Hemispheric loading (Mtonnes)	Total downward flux at the surface (W m^{-2})	
			Exact	Parameterization
1	0	0	509.2	509.2
2	0.0125	3.2	482.5	490.4
3	0.0625	15.9	393.1	422.1
4	0.125	31.9	307.4	349.9
5	0.250	63.8	192.7	240.5
6	0.375	95.6	124.4	165.3
7	0.625	159.4	56.0	78.1
8	1.25	318.8	11.0	12.0

The column densities which were assumed and the corresponding hemispheric loadings, followed by the total (direct plus diffuse) downward broad-band solar flux at the surface (exact) and the value given by a simple parameterization (see text).

would be expected to be minimal. It has been demonstrated that heating of the smoke could lead to condensation and hence to interaction between cloud water and smoke particles.

Received 5 August; accepted 31 October 1985.

1. Turco, R. P., Toon, O. B., Ackerman, T. P., Pollack, J. B. & Sagan, C. *Science* **222**, 1283–1292 (1983).
2. Crutzen, P. J., Galbally, I. E. & Bruhl, C. *Climatic Change* **6**, 323–634 (1984).
3. Covey, C., Schneider, S. H. & Thompson, S. L. *Nature* **308**, 21–25 (1984).
4. Alexandrov, V. V. & Stenichikov, G. L. in *Proc. Appl. Math.*, Moscow (1983).
5. *The Effects on the Atmosphere of a Major Nuclear Exchange* (National Academy Press, Washington, 1985).
6. Thompson, S. L. *Nature* **317**, 35–39 (1985).
7. Ramaswamy, V. & Kiehl, J. T. *J. geophys. Res.* **90**, 5597–5613 (1985).
8. Tapp, M. C. & White, P. W. *Q. J. R. met. Soc.* **102**, 277–296 (1976).
9. Carpenter, K. M. *Q. J. R. met. Soc.* **105**, 629–688 (1979).
10. Golding, B. W. & Machin, N. A. *Proc. Nowcasting-II Symp.*, Norrköping (ESA-SP 208, 1984).
11. Slingo, A. & Schrecker, H. M. *Q. J. R. met. Soc.* **108**, 407–426 (1982).
12. McClatchey, R. A., Fenn, R. W., Selby, J. E. A., Volz, F. E. & Garing, J. S. *Envir. Res. Pap.* No. 411 (Air Force Cambridge Research Laboratories, Massachusetts, 1972).

Evolution of an impact-induced atmosphere and magma ocean on the accreting Earth

Takafumi Matsui & Yutaka Abe

Geophysical Institute, Faculty of Science, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Early rapid formation of the atmosphere and hydrosphere on the terrestrial planets has recently been proposed^{1,2}. Here we present a quantitative study of this process during accretion by planetesimal impacts. These impacts increase the surface temperature and thus affect the formation of either a proto-atmosphere or a proto-hydrosphere by degassing of volatiles. We show that an impact-induced H₂O atmosphere increases the surface temperature of the Earth to a stage where a magma ocean is possible, with the total amount of H₂O in the proto-atmosphere clustering around 10²¹ kg, this number is rather insensitive to variations in the input data. We suggest that the apparent coincidence of the H₂O abundance in the proto-atmosphere with the present mass (~1.4 × 10²¹ kg) of the ocean is evidence for an impact origin for the atmosphere and hydrosphere.

Planets were formed by accretion of planetesimals in heliocentric orbits³. High-velocity planetesimal impacts produce craters on the surface of a growing planet. Cratering includes not only the formation of a cavity and the release of impact energy but also the ejection of dust and the shock-heated volatile gases, such as H₂O and CO₂, retained in planetesimals and the surface layer. The frequency of high-velocity collisions during planetary accretion results in an impact-induced atmosphere surrounding the entire surface of a growing planet. The existence of such an atmosphere affects the energy balance between the impact energy released in the surface layer, radiative heat transfer from the surface to interplanetary space, and heating of the surface layer. Previous studies have ignored this effect and assumed a fixed surface temperature^{4,5}. A more complete description of the early thermal history of a planet needs to include simultaneous growth by planetesimal impacts and the evolution of an impact-generated atmosphere. As H₂O is the most abundant volatile on the Earth, we consider an evolution of the atmosphere consisting only of H₂O in the following analysis.

The formation of an H₂O atmosphere depends on the surface temperature T_s . When $T_s \leq 900$ K, impact dehydration is a promising source mechanism to produce such an atmosphere. According to Lange and Ahrens⁶, serpentine loses its structural water when peak impact pressure exceeds the critical impact pressure required for the dehydration reaction of serpentine. The critical impact pressure is estimated to be $P_{cr} = 6 \times 10^{10}$ Pa when the surface layer and planetesimals have zero porosity⁶.

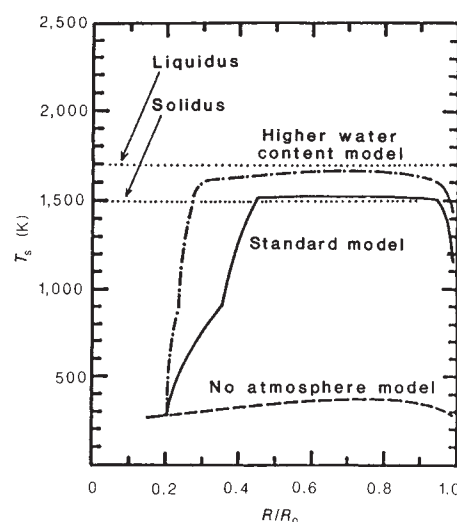


Fig. 1 The evolution of surface temperature during accretion for the standard Earth and higher W_{prj} models. The radius is normalized by the final value. The broken line traces the surface temperature of the model without an impact-generated atmosphere ($\tau_{acc} = 5 \times 10^7$ yr). The surface temperature begins to increase once the impact velocity exceeds a critical value. The rapid rise in the surface temperature after the Earth grows to $\sim 0.3 R_0$ is due to an increase in the total mass of the atmosphere because of initiation of the complete dehydration reaction of the surface layer. Once the surface temperature reaches the melting temperature, it becomes constant.

P_{cr} is a function of the porosity; P_{cr} reduces to 2.28×10^{10} Pa for a porosity of 17%. Because the surface of the growing Earth was heavily bombarded by high-velocity planetesimal impacts, we used $P_{cr} = 2.28 \times 10^{10}$ Pa in our model. The peak impact pressure is given by $P_{imp} = \rho[C_0 + (K' + 1)V_i/8]V_i/2$, where ρ , C_0 and K' are density, bulk sound velocity and partial derivative of incompressibility with respect to pressure. The impact velocity is given by $V_i^2 = (2GM/R) + u^2$, where G , M , R and u are the gravitational constant, mass and radius of a growing Earth, and velocity of planetesimals relative to a growing Earth at very great distance ($u = (GM/R\theta)^{1/2}$, where θ is the Safronov number). We used $\rho = 4,200$ kg m⁻³, $C_0 = 3,000$ m s⁻¹, $K' = 5$ and $\theta = 4$ for the standard model.

Once the peak pressure exceeds the critical impact pressure, the dehydration reaction is assumed to proceed completely. Then the growth of the impact-generated H₂O atmosphere during the Earth's accretion from M to $M + \Delta M$ can be expressed by

$$\Delta M_a = \Delta M W_{prj} f$$

where ΔM_a is the increase in mass of the H₂O atmosphere, W_{prj} is the water content of the planetesimals and f = (amount of the water added to the atmosphere)/(amount of the water released by shock heating). If all the water added to the Earth is now in the hydrosphere, $W_{prj} = 0.027\%$. This is a minimum estimate of W_{prj} and we used $W_{prj} = 0.1\%$ in our standard model. It is, however, difficult to estimate both W_{prj} and f , the latter ratio primarily expressing the loss mechanism of H₂O. The water vapour may be removed from the atmosphere by the following processes: (1) dissociation of H₂O into hydrogen and oxygen by photochemical reactions, with subsequent escape of hydrogen into space; (2) thermal escape of water vapour into interplanetary space without dissociation; (3) oxidation of the surface aided by photostimulation; (4) re-hydration with the surface minerals; (5) oxidation of Fe metal; and (6) dissociation of water into silicate melt. Processes (1)–(3) are not important because the accretion time required for the growth of ΔM is much shorter than the characteristic decay time of those mechanisms^{6,7}. Mechanism (5) is also not important⁷. Therefore, the

hydration reaction may be the most dominant loss mechanism in this temperature range. As the reaction rate of the hydration reaction is fast compared with the accretion time, the loss of H_2O may be limited by the amount of available reactants: that is, the thickness of the layer wherein the hydration reaction can occur determines the loss of H_2O . Shock dehydration occurs even in relatively deep regions, but dehydration reactions occur only in the uppermost layer which is in contact with the atmosphere for a long time. The amount of released water is qualitatively larger than that of rehydrated water, and hence an atmosphere may be formed. We adopted $f = 0.2$ as the standard model. In fact, the choice of f does not affect the evolution of the atmosphere because the surface temperature reaches 900 K even for a very small value.

When T_s is > 900 K and less than the melting temperature T_m , hydrous minerals become unstable. Then, instead of a hydration reaction, a dehydration reaction occurs even without impact heating. Therefore, we can consider that $f = 1$ in this temperature range.

When $T_s \geq T_m$, the sixth mechanism dominates the loss mechanism for H_2O because H_2O dissolves into a silicate melt under high atmospheric pressure. According to Fricker and Reynolds⁸, the amount X_w (in weight %) of water dissolved into silicate melt at pressure, P_a (in pascal), is approximately given by

$$X_w = 2.08 \times 10^{-6} P_a^{0.54}$$

Assuming a plane-parallel atmosphere, $P_a = M_a g / 4\pi R^2$, where M_a and g are the total mass of the H_2O atmosphere and the gravity at the surface of a growing Earth, respectively. The growth of the atmosphere in this temperature range can be simply calculated by

$$\Delta M_a = \Delta M (W_{prj} - \alpha X_w)$$

where α is the degree of melt determined from the distribution of the latent heat between the solidus and liquidus. The solidus and liquidus temperatures are assumed to be $T_{sol} = 1,500$ K and $T_{liq} = T_{sol} + 200$, respectively. The latent heat (400 kJ kg^{-1} K $^{-1}$) is assumed to be uniformly distributed between solidus and liquidus temperatures.

Next, we estimated the effect of the H_2O atmosphere on the surface temperature of the Earth growing by planetesimal impacts. The energy balance equation at the surface layer of a growing Earth is given approximately by

$$h \Delta M V_i^2 / 2 = 4\pi R^2 F_{atm} \Delta t + C_p \Delta M (T_s - T_p)$$

where h , Δt , C_p , T_s and T_p are the fraction of the released impact energy retained in the surface layer, the time interval required for the growth from M to $M + \Delta M$, the specific heat of the surface layer, the temperature of the surface layer and the original temperature of the accreting planetesimal⁷. To derive this equation, for simplicity, we assumed impact energy as the primary heat source and a δ -function-like size-distribution of planetesimals. The quantity h is, in fact, close to but not exactly 1, which means that some fraction of impact energy was imparted into the sub-surface layers; for simplicity, however, we take $h = 1$. However, this assumption does not affect the estimate of the surface temperature because $T_s \propto h^{1/4}$. We used $C_p = 10^3$ J kg^{-1} K $^{-1}$, and $T_p = 250$ K as the standard model. F_{atm} is the energy flux escaping from the surface into interplanetary space through the atmosphere:

$$F_{atm} = 2\sigma(T_s^4 - T_0^4)/(\tau_s^* + 2)$$

where σ and T_0 are the Stefan-Boltzmann constant and the black body equilibrium temperature of the proto-Earth heated only by solar irradiation, and τ_s^* corresponds approximately to an optical depth for the optically thin atmosphere. We assumed that the atmosphere is gray, plane-parallel and radiatively equilibrated. Then, τ_s^* can be defined by $\tau_s^* = (3/2) \int_0^\infty \kappa \rho_a dz = (3\kappa M_a)/(8\pi R^2)$, where κ is the coefficient of absorption⁷.

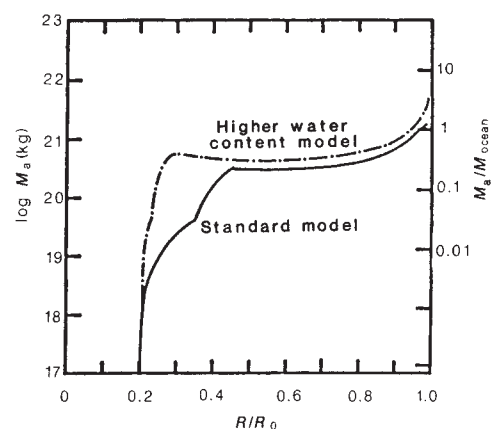


Fig. 2 The total mass of the impact-generated H_2O atmosphere is plotted against the normalized radius for the standard ($W_{prj} = 0.1\%$) Earth and higher water-content ($W_{prj} = 1\%$) models. Note that the total mass remains nearly constant after the Earth grows to $0.4 R_0$ and is very close to the present mass of the Earth's oceans ($\sim 1.4 \times 10^{21}$ kg).

Assuming hydrostatic equilibrium and direct proportion of κ to pressure, κ is given by $\kappa = (\kappa_0 g / 3P_0)^{1/2}$, where κ_0 is the coefficient of absorption at pressure P_0 (ref. 7). We should use an absorption coefficient averaged over a wide range of wavelengths because we assumed a gray atmosphere. We adopted a value of the smoothed absorption coefficient around the window region ($\sim 1,000$ cm^{-1}) as κ_0 : $\kappa_0 = 0.01$ m 2 kg $^{-1}$ and $P_0 = 101,325$ Pa (ref. 9). Note that the κ_0 value thus determined is a minimum. As the mass of the atmosphere increases, the lower atmosphere becomes optically thick and equilibrates convectively. A straightforward estimate of the surface temperature follows from assuming that $T_s = (P_s / P_{eff})^{(\gamma-1)/\gamma} T_{eff} = (\tau_s^*)^{(\gamma-1)/\gamma} T_{eff}$, where T_{eff} is the effective temperature given by $T_{eff} = (F_{atm} / \sigma + T_0^4)^{1/4}$, P_{eff} and P_s are the atmospheric pressure at the tropopause and the base of the atmosphere, and γ is the ratio between the specific heat with constant pressure and the specific heat with constant volume ($= C_p / C_v$). In this case, $F_{atm} = (T_s^4 / \tau_s^* - T_0^4)$. This gives an energy flux that is lower than the flux given for the radiative case, thereby resulting in a higher surface temperature. However, the difference is small. For simplicity, we use the energy flux equation for the radiative case to calculate the surface temperature even for an optically thick atmosphere.

We need an accretion model to solve the energy balance equation. According to Safronov³, the accretion rate of the Earth is given by

$$\Delta M / \Delta t = 4\pi R^2 (1 + 2\theta) \sigma_p / P_k$$

where P_k is the period of Kepler motion of the accreting Earth and σ_p is the surface density. We simply assumed $\sigma_p = (M_0 - M) / S$, where M_0 and S are the final mass of the Earth and the area from which the Earth collects planetesimals. We used $M_0 = 6.0 \times 10^{24}$ kg and $S = 4.08 \times 10^{22}$ m 2 , which gives an accretion time for the Earth of about 5×10^7 yr.

By integrating numerically the above equations with respect to time we determine the evolution of the surface temperature and impact-generated H_2O atmosphere. Figure 1 shows a trace of the surface temperature of the standard Earth and higher W_{prj} (1%) model. The surface temperature increases due to the blanketing effect of the atmosphere, and the surface of the accreting Earth begins to melt once the radius exceeds $0.4 R_0$, where R_0 is the final radius ($R_0 = (3M_0 / 4\pi\rho)^{1/3} = 6,987$ km). After the surface temperature reaches the melting temperature, it becomes nearly fixed at about the melting temperature. This is because an increase in the surface temperature increases the degree of melting, which causes efficient absorption of H_2O .

molecules from the atmosphere. A decrease in the H_2O content of the atmosphere results in decreasing the efficiency of the blanketing effect, thereby decreasing the surface temperature. A decrease in the surface temperature allows the surface to resolidify, but this increases the H_2O in the atmosphere. The surface temperature decreases rapidly at the final stage of growth. This is because the accretion rate assumed in this study decreases rapidly as the final radius is approached. The surface temperature, however, may not decrease in the final stages when the effect of solar irradiation on the atmosphere (the so-called greenhouse effect) is taken into account. The evolution of the atmosphere after solar irradiation becomes a primary heat source (replacing impact energy) is one of the most important problems for future study.

Figure 2 shows the evolution of the H_2O atmosphere for the standard Earth model and for a model with higher water content ($W_{\text{prj}} = 1\%$, with other parameters the same as the standard model). The amount of the H_2O atmosphere becomes nearly constant once the radius exceeds about $0.4 R_0$, at which point the surface temperature reaches the melting temperature (see Fig. 1). This confirms the above interpretation concerning the constant surface temperature, where the solubility of water into silicate melt controls the amount of water in the atmosphere. The atmospheric pressure and τ_s^* also become nearly constant (at $\sim 10^7$ Pa and 10^3 , respectively). It is shown elsewhere¹⁰ that melting the surface layer requires an atmosphere whose τ_s^* is 10^3 . At the final stages of growth the total mass of the atmosphere increases gradually. This is caused by resolidification of the surface layer.

Comparison of the standard Earth and higher water-content models reveals that the final mass of H_2O atmosphere does not differ significantly between these models even though the time at which the mass of the atmosphere reaches this value differs. This suggests that the amount of H_2O in the atmosphere does not depend on the water content of the raw materials; rather it depends on the surface temperature during accretion. The total interior water content, however, is plainly dependent on the original water content. The surface temperature history for the higher water-content model also resembles that shown in Fig. 1. The initial thermal profile is shown not to depend heavily on the water content. It is very interesting to note that the final mass of the H_2O atmosphere is fairly close to the present mass of the oceans ($\sim 1.4 \times 10^{21}$ kg). The level at which the mass of the H_2O atmosphere becomes nearly constant is only weakly dependent on the accretion time and absorption coefficient. The depth of a magma ocean and degree of partial melt also depend on these parameters. For W_{prj} greater than 1%, the level of constant mass of the H_2O atmosphere depends weakly on this quantity. This is because the solubility of water in silicate melt at $\sim 10^7$ Pa is $\sim 1\%$. A choice of reasonable values for these and other parameters, as shown here, leads to an apparent coincidence of the H_2O abundance in the proto-atmosphere with the present mass of the oceans. This suggests that an impact origin for the atmosphere and hydrosphere is feasible. Incorporating CO_2 would increase the optical depth by several hundred, but this increase does not significantly change the conclusion derived in this study. The oldest rock ($\sim 3.8 \times 10^{-1}$ yr) known so far is of sedimentary origin, which suggests that an extensive hydrosphere was already formed at that time. We conclude that the present mass of the oceans provides evidence for their impact origin and for the past existence of a magma ocean on the proto-Earth.

We thank P. H. Schultz, T. J. Ahrens, and S. C. Solomon for their critical comments on an earlier draft of this paper. This research was partially supported by grants-in-aid for Scientific Research (Nos 60540247 and 59390007) sponsored by the Ministry of Education, Science, and Culture of Japan. The final stage of preparation of this work was carried out while T.M. was on leave at the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology.

Received 3 June; accepted 20 November 1985.

1. Fanale, F. P. *Chem. Geol.* **8**, 79–105 (1971).
2. Cogley, J. G. & Henderson-Sellers, A. *Rev. Geophys. space Phys.* **22**, 131–175 (1984).
3. Safronov, V. S. *NASA Techn. Transl.*, TTF-677, p. 206 (1972).
4. Safronov, V. S. *Icarus* **33**, 3–12 (1978).
5. Kaula, W. M. *J. geophys. Res.* **84**, 999–1088 (1979).
6. Lange, M. A. & Ahrens, T. J. *Icarus* **51**, 96–120 (1982).
7. Abe, Y. & Matsui, T. *J. geophys. Res.* **89**, C323–C328 (1984).
8. Fricker, P. E. & Reynolds, R. T. *Icarus* **9**, 221–230 (1968).
9. Yamamoto, G. *Sci. Rep. Tohoku Univ. Ser. 5* **4**, 9–23 (1952).
10. Matsui, T. & Abe, Y. *Earth, Moon Planets* (in the press).

Chemical effects of large impacts on the Earth's primitive atmosphere

Bruce Fegley Jr, Ronald G. Prinn,
Hyman Hartman & G. Hampton Watkins*

Department of Earth, Atmospheric, and Planetary Sciences,
Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, USA

Intense bombardment of the moon and terrestrial planets ~ 3.9 – 4.0×10^9 years ago^{1,2} could have caused the chemical reprocessing of the Earth's primitive atmosphere³. In particular, the shock heating and rapid quenching caused by the impact of large bodies into the atmosphere could produce molecules such as HCN and H_2CO ⁴ which are important precursors for the abiotic synthesis of complex organic molecules^{5–7}. Here we model the production of HCN and H_2CO by thermochemical equilibrium and chemical kinetic calculations of the composition of shocked air parcels for a wide range of temperatures, pressures and initial compositions. For atmospheres with $\text{C/O} \geq 1$, our results suggest that bolide impacts cause HCN volume mixing ratios of approximately 10^{-3} to 10^{-5} in the impact region and global average ratios of 10^{-5} to 10^{-12} . The corresponding H_2CO mixing ratios in the impact region are 10^{-7} to 10^{-9} ; no-global mixing can occur, however, as H_2CO is rapidly destroyed or rained out of the atmosphere within days to hours. Rainout to the oceans of 3–15% of the HCN produced can provide $\sim (3\text{--}14) \times 10^{11}$ mol HCN per year. This is somewhat larger than other predicted sources of HCN⁸ and H_2CO ⁹ on the primitive Earth.

The atmosphere appeared on Earth very early in its history. Argon and xenon isotope geochemistry^{10,11} indicate that the mean age of the atmosphere is 4.4×10^9 yr. The carbonate-bearing sedimentary rocks in the Isua, Greenland supracrustal rocks further suggest the presence of an atmosphere-ocean system 3.8×10^9 yr ago¹². To evaluate the effects of impacts on this early atmosphere we consider a suite of more than 100 pressure (P) and temperature (T) composition points in the H–C–N–O tetrahedron ranging from chemically reducing to chemically neutral atmospheres, shock pressures from 10^{-3} to 10^3 bar, and shock temperatures from 500 to 5,000 K. Although precise knowledge of the composition of the pre-Archean atmosphere is unavailable, a plausible range of primitive atmosphere compositions can be defined by considering the oxidation states and compositions of present-day volcanic gases, of volcanic gases in equilibrium with metallic iron and of the outgassing from chondritic planets^{12–15}. Plausible atmospheres all contain N_2 and small amounts of H_2O in equilibrium with liquid water. Carbon may occur predominantly as CH_4 , CO, or CO_2 depending on the assumed oxidation state and H_2 may be present in small amounts conversant with its rapid atmospheric escape.

Thermochemical equilibrium calculations were done using a Gibbs free energy minimization code¹⁶; thermodynamic data for more than 50 compounds in the H–C–N–O system were taken from standard sources^{17,18}. The quenched abundances of HCN

* Present address: J. S. Lewis Associates Inc., Tucson, Arizona 85745, USA.

Table 1 Reactions and rate constants used to estimate t_{chem} values for HCN and H_2CO

Reaction	Rate constant	Source
(1) $\text{CHO} + \text{H} \rightarrow \text{CO} + \text{H}_2$	$6 \times 10^{13} \exp(-2,500/T)$	19
(2) $\text{CHO} + \text{M} \rightarrow \text{CO} + \text{H} + \text{M}$	$7 \times 10^{13} \exp(-7,550/T)$	19*
	$2 \times 10^{13} T^{1/2} \exp(-14,000/T)$	
(3) $\text{CHO} + \text{O} \rightarrow \text{CO} + \text{OH}$	3×10^{13}	19
(4) $\text{CHO} + \text{O} \rightarrow \text{CO}_2 + \text{H}$	3×10^{13}	19
(5) $\text{CHO} + \text{OH} \rightarrow \text{CO} + \text{H}_2\text{O}$	6×10^{12}	19*
	3×10^{13}	
(6) $\text{CHO} + \text{O}_2 \rightarrow \text{CO} + \text{HO}_2$	3×10^{13}	19*
	$6 \times 10^{12} \exp(-3,650/T)$	
(7) $\text{H}_2\text{CO} + \text{H}_2 \rightarrow \text{OH} + \text{CH}_3$	$1.4 \times 10^{14} \exp(-36,200/T)$	21
(8) $\text{H}_2\text{CO} + \text{M} \rightarrow \text{HCO} + \text{H} + \text{M}$	$4 \times 10^{12} \exp(-18,500/T)$	22†
(9) $\text{CN} + \text{NO} \rightarrow \text{CO} + \text{N}_2$	8.9×10^{12}	20‡
(10) $\text{CN} + \text{NO} \rightarrow \text{NCO} + \text{N}$	$1.0 \times 10^{14} \exp(-21,190/T)$	23
(11) $\text{CN} + \text{CN} \rightarrow \text{C}_2 + \text{N}_2$	$2.5 \times 10^{17} \exp(-48,000/T)$	20*
	$1.6 \times 10^{15} \exp(-21,700/T)$	
(12) $\text{CN} + \text{O} \rightarrow \text{CO} + \text{N}$	1.0×10^{13}	20
(13) $\text{CN} + \text{O}_2 \rightarrow \text{CNO} + \text{O}$	$2.4 \times 10^{13} \exp(-450/T)$	20
(14) $\text{CN} + \text{C} \rightarrow \text{N} + \text{C}_2$	$3.0 \times 10^{14} \exp(-18,000/T)$	20
(15) $\text{CN} + \text{N} \rightarrow \text{C} + \text{N}_2$	$4.4 \times 10^{14} \exp(-4,600/T)$	20
(16) $\text{HCN} + \text{H}_2 \rightarrow \text{CH}_2 + \text{NH}$	$6.5 \times 10^{15} \exp(-70,456/T)$	24
(17) $\text{HCN} + \text{M} \rightarrow \text{H} + \text{CN} + \text{M}$	$1.0 \times 10^{16} \exp(-54,650/T)$	25
(18) $\text{HCN} + \text{O} \rightarrow \text{NCO} + \text{H}$	$7.2 \times 10^{13} \exp(-7505/T)$	26

Units for two-body rate constants are $\text{cm}^3 \text{mol}^{-1} \text{s}^{-1}$.

* Calculations done with both expressions.

† Also see Bowman²⁷ for a similar expression.

‡ Average of values given by Baulch *et al.*²⁰.

and H_2CO in the shock wave were estimated using the relationship

$$t_{\text{chem}}(T_Q) \approx t_{\text{cool}}(T_Q) \quad (1)$$

where t_{chem} is the chemical lifetime of HCN or H_2CO , t_{cool} is the characteristic cooling time of the shock wave and T_Q is the quench temperature. The t_{chem} values for HCN and H_2CO were estimated using kinetic data for the relevant destruction reactions of CN and HCO (see Table 1)^{19–27}. This treatment is valid when HCN and CN, and H_2CO and HCO remain in equilibrium (which we verified). Analogous calculations have also been done by previous investigators estimating NO_x production from high temperature shock phenomena (such as nuclear explosions^{28–32}, lightning discharges^{8,33} and large impacts in the atmosphere^{34,35}).

The calculated thermochemical equilibrium mixing ratios of important species in the shocked air parcels are illustrated in Fig. 1. These profiles are representative of those obtained by shocking atmospheric compositions which are chemically neutral (Fig. 1a) or chemically reducing (Fig. 1b). The major species produced from a CO_2 -rich atmosphere are CO, O_2 , H_2 and NO; CO-rich atmospheres yield primarily CO_2 , H_2 , CH_4 , HCN, NH_3 and H_2CO . Graphite precipitation is also possible in CO-rich systems but may be kinetically inhibited if nucleation is slow relative to t_{cool} .

Calculations over a wide range of assumed atmospheric compositions show that the C/O atomic ratio strongly influences the yields of the shock products (see Table 2). Our calculated HCN and NO abundances follow the same pattern first noted by Chameides and Walker⁸. We find that the interconversions between HCN and NO for different major carbon-bearing species are described by the reversible equilibrium reactions:



The thermochemical equilibrium calculations also show that the HCN volume mixing ratio near T_Q is not sensitive to large changes in the total pressure. It decreases by only two orders of magnitude as P varies from 10^3 to 10^{-3} bar. However, the

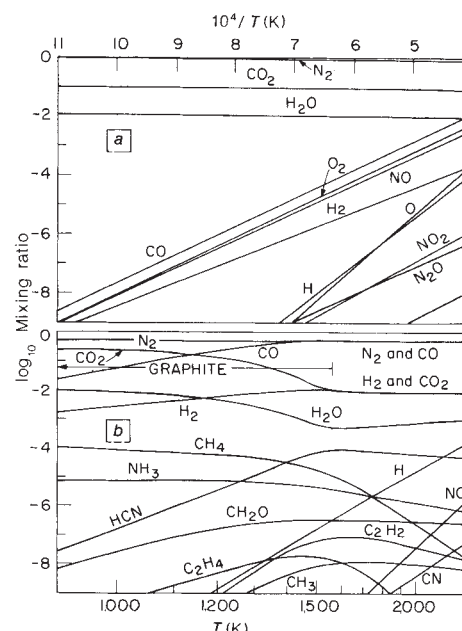


Fig. 1 Equilibrium abundances of important species in a cooling parcel of shocked air. *a*, Neutral atmosphere of 90% N_2 , 9% CO_2 , 1% H_2O at a shock pressure of 10 bar. HCN, H_2CO , NH_3 and many other species with mixing ratios $< 10^{-9}$ are not shown here. *b*, Reducing atmosphere of 49.5% N_2 , 49.5% CO , 1% H_2O at a shock pressure of 100 bar. Graphite precipitation occurs at 1,556 K; it is a major carbon sink by 1,000 K. All hydrocarbons (except C_2H_2 and C_2H_4) have mixing ratios $< 10^{-9}$. In both cases increasing dissociation of molecules to radicals, atoms, and ions occurs with increasing temperature.

H_2CO mixing ratio varies as P above the graphite condensation temperature and is pressure independent below this point.

Calculated chemical lifetimes for HCN and H_2CO are compared with the radiative cooling time for the shocked air column in Fig. 2. The HCN quench temperatures vary from 1,560 to 2,100 K depending on the specific HCN destruction reactions. The H_2CO quench temperatures are in contrast $< 1,110$ K. If reaction (2) in Table 1 is relatively fast, H_2CO may remain in equilibrium below 650 K. The corresponding mixing ratios for HCN ($T_Q \approx 1,600$ K) and for H_2CO ($T_Q \approx 1,000$ K) from Fig. 1b are 10^{-4} and 10^{-8} , respectively. For the range of reducing atmospheres studied we calculate HCN mixing ratios of 10^{-3} to 10^{-5} at T_Q and H_2CO mixing ratios of $\sim 10^{-7}$ to 10^{-9} at T_Q in the impact region.

Globally averaged HCN mixing ratios were estimated using a simple model developed to describe NO_x production by the putative Cretaceous/Tertiary impactor³⁴. We take the HCN production efficiency as 10^7 molecules erg^{-1} (for a $\text{CO} + \text{N}_2$ atmosphere) from Chameides and Walker⁸. This efficiency is strongly composition-dependent ranging from $\sim 10^{10}$ molecules erg^{-1} ($\text{CH}_4 + \text{N}_2$) to $\sim 10^3$ molecules erg^{-1} ($\text{CO}_2 + \text{N}_2 + \text{H}_2\text{O}$)⁸. From Lewis *et al.*³⁴ we take the energy deposited in the atmosphere by a 10^{17} g metal-rich impactor (vertical entry angle ~ 12 kms s^{-1}) as $\sim 10^{25.8}$ erg and the energy deposition by a 10^{18} g cometary impactor (grazing entry angle ~ 65 km s^{-1}) as $\sim 10^{30.3}$ erg. The estimated HCN yields are in the range of 10^{33} – 10^{37} molecules per impact and the fraction of the atmosphere shocked per impact is in the range 10^{-8} – 10^{-3} (assuming HCN mixing ratios of $\sim 10^{-3}$ to 10^{-5} at T_Q and the present atmospheric mass of 5.1×10^{21} g). Corresponding globally-averaged HCN and H_2CO mixing ratios would be from 10^{-5} to 10^{-12} and from 10^{-10} to 10^{-16} , respectively. However, these globally-averaged values are relevant only if the chemical lifetimes of the species are longer than the global atmospheric mixing time of 1–3 yr.

Several potential loss mechanisms for HCN and H_2CO are

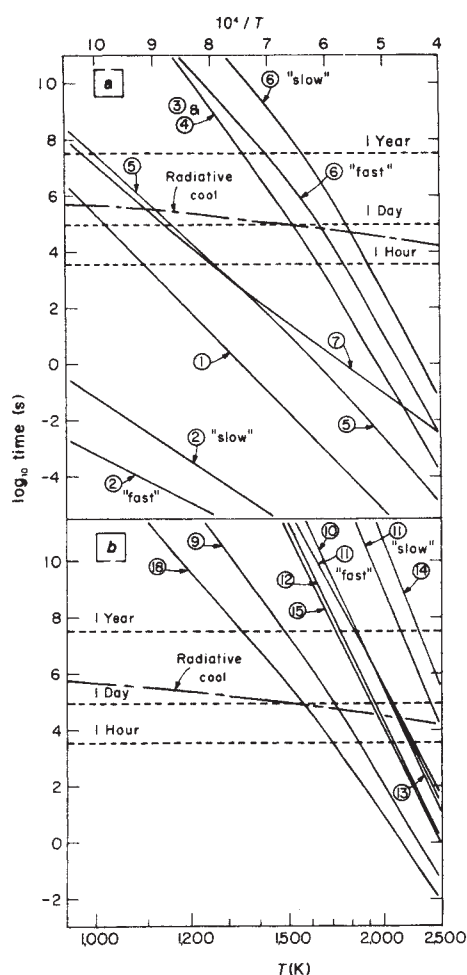


Fig. 2 Calculated chemical lifetimes for H_2CO (a) and HCN (b) for the reducing atmosphere shown in Fig. 1b. The numbers refer to the reactions in Table 1; reaction 16 has a t_{chem} very similar to those for reactions 12, 13, 15 and is not shown. The radiative cooling time is calculated from the Stefan-Boltzman law. For $T > T_Q$, $t_{\text{chem}} < t_{\text{cool}}$ and the HCN and H_2CO remain in equilibrium, while for $T < T_Q$, $t_{\text{chem}} > t_{\text{cool}}$ and the HCN and H_2CO are quenched at their mixing ratios established at $T = T_Q$ where $t_{\text{chem}} = t_{\text{cool}}$. The T_Q values used for HCN and H_2CO in our discussion are 1,600 K and 1,000 K respectively. However, H_2CO may remain in equilibrium below ~ 650 K if reaction (2) is fast.

possible; we qualitatively discuss these by analogy with these processes in the present-day terrestrial atmosphere. Precipitation can transport HCN and H_2CO to the primitive oceans where abiotic syntheses of complex organic molecules may occur^{5-7,12}. Cicerone and Zellner³⁶ deduced that HCN has a long rainout lifetime of about 34 yr; H_2CO is much more soluble (it polymerizes in aqueous solution) and has a rainout lifetime³⁷ of about 10 days. Both species are also destroyed by solar ultraviolet photolysis and by reaction with OH . The HCN and H_2CO chemical lifetimes in the present day atmosphere are ~ 1 –5 years³⁷ and ~ 3 hours³⁸, respectively. However, the latter value refers to the sunlit troposphere at 30°N and at noon; a diurnal average is probably 3–4 times longer. These general considerations suggest that isolated impacts are a negligible global H_2CO source relative to possible photolytic sources⁹; only small, temporally and spatially localized production may occur. However, these considerations are consistent with a significant global source of HCN from isolated impacts.

To a first order, $1/34$ to $5/34$ of the HCN produced in a single impact event will be rained out of the atmosphere based on the lifetimes quoted above. Taking the present globally averaged rainfall rate³⁷ of $3.3 \times 10^{-6} \text{ g cm}^{-2} \text{ s}^{-1}$, a Henry's law constant³⁶

Table 2 Equilibrium mixing ratios of HCN , NO and H_2CO for different assumed atmosphere compositions at 2,000 K and shock pressures of 10 and 100 bar

Starting composition	$p = 10 \text{ bar}$		
	$\log_{10} \text{ mixing ratios}$		
	HCN	NO	H_2CO
44.4% N_2 , 44.4% CO , 11.1% H_2 *†	-2.6	-9.3	-6.6
14.3% N_2 , 57.1% CO , 22.6% CH_4 †	-2.6	-9.6	-6.0
9% N_2 , 90.1% CO , 1% H_2O	-5.0	-7.5	-7.2
49.5% N_2 , 49.5% CO , 1% H_2O	-5.2	-7.0	-7.5
1% N_2 , 98% CO , 1% H_2O	-5.4	-8.1	-7.2
70% N_2 , 26.4% CO_2 , 2.6% CO , 1% H_2O	-9.9	-4.1	-10.4
80% N_2 , 17.3% CO_2 , 1.7% CO , 1% H_2O	-10.0	-4.1	-10.6
80% N_2 , 19% CO_2 , 1% H_2O	-12.2	-3.2	-12.3

Starting composition	$p = 100 \text{ bar}$		
	$\log_{10} \text{ mixing ratios}$		
	HCN	NO	H_2CO
10% N_2 , 80% CO , 10% H_2 *	-3.2	-8.8	-6.3
9% N_2 , 90% CO , 1% H_2O	-4.0	-8.0	-6.2
49.5% N_2 , 49.5% CO , 1% H_2O	-4.1	-7.4	-6.5
1% N_2 , 98% CO , 1% H_2O	-4.4	-8.5	-6.2
70% N_2 , 26.4% CO_2 , 2.6% CO , 1% H_2O	-8.9	-4.6	-9.4
80% N_2 , 17.3% CO_2 , 1.7% CO , 1% H_2O	-9.0	-4.6	-9.6
80% N_2 , 19% CO_2 , 1% H_2O	-12.0	-3.4	-11.9

* High H_2 mixing ratio cannot be sustained for long periods. This case is shown for illustrative purposes only.

† Graphite saturated system. This case is shown for illustrative purposes only.

of 4×10^3 and a HCN global mixing ratio of 10^{-6} we derive a HCN rainout rate of 3 to $14 \times 10^{11} \text{ mol yr}^{-1}$. This is somewhat larger than predicted H_2CO (from photolysis)⁹ and HCN (from lightning)⁸ rainout rates on the primitive Earth of $\sim 10^{11} \text{ mol yr}^{-1}$.

Impactor fluxes on the primitive Earth are not well constrained. However, a few large impacts every 1–5 yr will maintain the globally averaged HCN mixing ratios and rainout rates calculated above. More frequent impacts will lead to increasing HCN abundances.

Preliminary stages of this work were supported by NASA grant NGR-22-007-269 to Harvard University. The final part was supported by NSF grants ATM-84-01232, NAG9-108 and NAGW-114 to MIT.

Received 23 September; accepted 29 November 1985.

1. Tera, F., Papanastassiou, D. A. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **22**, 1–21 (1974).
2. Wetherill, G. W. *Proc. 6th lunar Sci. Conf.* 1539–1561 (1975).
3. Hochstim, A. R. *Proc. natn. Acad. Sci. U.S.A.* **50**, 200–208 (1963); in *Chemical Evolution and the Origin of Life* (eds Buvet, R. & Ponnampemura, C.) 96–113 (North-Holland, New York, 1971).
4. Fegley, B., Hartman, H. & Prinn, R. G. *EOS* **63**, 1018 (1982).
5. Oro, J. & Kimball, A. P. *Archs biochem. Biophys.* **94**, 217–227 (1961).
6. Cairns-Smith, A. G. *Genetic Takeover and the Mineral Origins of Life*, 21–31 (Cambridge University Press, 1982).
7. Abelson, P. H. *Proc. natn. Acad. Sci. U.S.A.* **1365**–1372 (1966).
8. Chameides, W. L. & Walker, J. C. G. *Origins Life* **11**, 291–302 (1981).
9. Pinto, J. P., Gladstone, G. R. & Yung, Y. L. *Science* **210**, 183–185 (1980).
10. Sarda, P., Staudacher, Th. & Allegre, C. J. *Earth planet. Sci. Lett.* **72**, 357–375 (1985).
11. Staudacher, Th. & Allegre, C. J. *Earth planet. Sci. Lett.* **60**, 389–406 (1982).
12. Holland, H. D. *The Chemical Evolution of the Atmosphere and Oceans* (Princeton University Press, 1984).
13. Holland, H. D. in *Petrologic Studies: A Volume to Honor A. F. Buddington* (ed Engel, A. E. J., et al.) 447–477 (Geological Society of America, Boulder, 1962).
14. Gerlach, T. M. & Nordlie, B. E. *Am. J. Sci.* **275**, 353–410 (1975).
15. Lewis, J. S. & Prinn, R. G. *Planets and their Atmospheres Origin and Evolution*, 225–238 (Academic, New York, 1984).
16. Gordon, S. & McBride, B. J. *NASA SP273* (1976).
17. *JANAF Thermochemical Tables* (eds Stull, D. R. & Prophet, H.) NSRDS-NBS-37 (US Government Printing Office, Washington, 1971).
18. Glushko, V. P. et al. (eds) *Thermodynamic Properties of Individual Substances Vols 1–4* (High-Temperature Institute, Moscow, 1978–1982).
19. Baulch, D. L., Drysdale, D. D., Duxbury, J. & Grant, S. J. *Evaluated Kinetic Data for High Temperature Reactions Vol. 3*, 337–361 (Butterworths, London, 1976).
20. Baulch, D. L., Duxbury, J., Grant, S. J., Montague, D. C. *Evaluated Kinetic Data for High Temperature Reactions Vol. 4*, 576–635 (*J. phys. Chem. Ref. Data* **10**, Suppl. 1, 1981).
21. Prinn, R. G. & Barshay, S. S. *Science* **198**, 1031–1034 (1977).
22. Chen, C. J. & McKenney, D. J. *Can. J. Chem.* **50**, 992–998 (1972).
23. Colket, M. B. *Int. J. Chem. Kinet.* **16**, 353–369.
24. Prinn, R. G. & Fegley, B. Jr. *Astrophys. J.* **249**, 308–317 (1981).
25. Szekely, A., Hanson, R. K., and Bowman, C. T. *J. phys. Chem.* **88**, 666–668 (1984).
26. Higashihara, T., Salto, K. & Murakami, I. *J. phys. Chem.* **87**, 3707–3712 (1983).
27. Bowman, C. T. *Symp. 15th Int. Combust. Proc.* 869–882 (Combustion Institute, Pittsburgh, 1974).
28. Foley, H. M. & Ruderman, M. A. *J. geophys. Res.* **78**, 4441–4450 (1973).
29. Gilmore, F. R. *J. geophys. Res.* **80**, 4553–4554 (1975).
30. Goldsmith, P., Tuck, H. F., Foot, J. S., Simmons, E. L. & Newson, R. L. *Nature* **244**, 545–551 (1973).

31. Johnston, H., Whitten, G. & Birks, J. *J. geophys. Res.* **78**, 6107–6135 (1973).
32. Zeldovich, Ya. B. & Raizer, Ya. *Physics of Shock Waves and High Temperature Phenomena* Vol. 2, 566–571 (Academic, New York, 1967).
33. Chameides, W. L. *Nature* **277**, 123–125 (1977).
34. Lewis, J. S., Watkins, G. H., Hartman, H. & Prinn, R. G. *Geol. Soc. Am. Spec. Pap.* **190**, 215–221 (1982).
35. Park, C. *Acta astronautica* **5**, 523–542 (1978).
36. Cicerone, R. J. & Zellner, R. *J. geophys. Res.* **88**, 10, 689–10, 696 (1983).
37. Giorgi, F. & Chameides, W. L. *J. geophys. Res.* **90**, 7872–7880 (1985).
38. Lowe, D. C. & Schmidt, U. *J. geophys. Res.* **88**, 10, 844–10, 858 (1983).

Direct conversion of methane to methanol, chloromethane and dichloromethane at room temperature

Kotaro Ogura & Kenji Takamagari

Department of Applied Chemistry, Yamaguchi University, Ube 755, Japan

As methane is extremely inert, its conversion into a form suitable for chemical use has been achieved by an indirect process in which methane is first converted to synthesis gas and then to methanol and to ethylene glycol¹. One disadvantage of this route of conversion is that it requires conditions of high temperature and pressure². Here we report the direct conversion of methane to methanol, chloromethane and dichloromethane at the three-phase interface (gas/solution/electrode) on illumination at room temperature. The key steps in this process are: (1) electrochemical oxidation of the chloride ion; (2) generation of the chlorine radical under illumination; and (3) formation of the methyl radical by the reaction of methane with the chlorine radical.

In the present study, we attempted the combined photochemical and electrochemical oxidation of methane at room temperature. A conventional two-compartment glass electrolysis cell was used, with the anode compartment separated from the cathode compartment by a fritted-glass partition. A water-sealed stopcock and a Luggin capillary provided contact between the anode and the reference electrode. The anode compartment incorporated a flat quartz window through which we illuminated the cell, using as a light source a 4-W low-pressure mercury lamp (Hamamatsu TV, L937-04). The wavelength of the illumination was 254 nm. The anode comprised a platinum plate partially immersed in the anolyte, and methane was introduced above the level of the solution. Only the surface of the platinum exposed to the gas phase was illuminated. To collect volatile products, the effluent gas was passed through a trap (25 cm³) containing 20% KOH by weight. Both the electrolytic solution and the trap were kept at 25 ± 1 °C. Various strengths of KCl solution were used as electrolytes; the results presented here are for a 0.6 M KCl solution of pH 11.0.

We determined the oxidation products present in both the electrolyte and residues in the trap by steam chromatography using an Ohkura Model SSC-1 which uses steam as the carrier gas and incorporates a flame ionization detector (FID) and a Poropak R column. This equipment is capable of detecting directly organic substances dissolved in an aqueous solution, without any pretreatment³.

The steam chromatograms of the solution after electrolysis at potentials more positive than +1.3 V (compared with the standard calomel electrode, SCE) showed four peaks at 2.0, 3.3, 10.0 and 26.1 min corresponding to chloromethane, methanol, dichloromethane and trichloromethane, respectively. However, the peak due to trichloromethane was very small compared with the other three peaks. No such peaks were observed for the solution electrolysed in the dark.

Table 1 shows the amounts of CH₃Cl, CH₃OH, CH₂Cl₂ and CHCl₃ produced in the 0.6 M KCl solution at pH 11.0 for various anodizing potentials. At potentials more negative than +1.0 V, there was no oxidation of methane. Conversion to CHCl₃ and

Table 1 Yield of chloromethane, methanol, dichloromethane and trichloromethane in a 0.6 M KCl solution at pH 11.0

V _e (V)	CH ₃ Cl	CH ₃ OH	CH ₂ Cl ₂	CHCl ₃
1.1	0.4	0.2	0.0	0.0
1.2	15.3	0.8	0.0	0.0
1.3	24.9	6.1	9.6	0.3
1.4	164.3	6.3	16.7	0.3
1.5	239.9	10.5	17.9	0.9
1.6	18.1	1.1	0.2	0.0
1.7	29.3	1.9	0.4	0.0
1.8	59.6	2.9	1.6	0.1
1.9	95.1	6.6	6.4	0.3
2.0	95.8	7.4	8.6	0.7

V_e, electrolysis potential compared with the saturated calomel electrode. Values shown are the quantities (in μmol) of electrolysis products. Electrolysis time was 3 h; volume of electrolyte 20 cm³; the anode comprised a platinum plate of area 6 cm².

CH₃OH commenced at +1.1 V, and further conversion to CH₂Cl₂ and CHCl₃ began at +1.3 V. At +1.5 V, the formation of these compounds was at a maximum, the current efficiencies being 12.8% (CH₃Cl), 0.38% (CH₃OH) and 2.05% (CH₂Cl₂) where the electric charge passed was 361.7 C. Chemical and gas-chromatographic analyses were done to detect other species such as HCHO or CO₂; however, the results were negative.

The results presented here indicate that the electrochemical evolution of the chlorine molecule and the formation of its radical on illumination are prerequisites for triggering the activation of methane. The photochemical chlorination of methane is well understood. Mixtures of methane and chlorine react rapidly in the presence of light. However, for this reaction to occur, the two gases must be mixed at an appropriate ratio, and the reaction is difficult to control as hydrogen atoms are replaced by chlorine atoms almost at random. In the combined photochemical/electrochemical method described here, the chlorination can be stopped after the initial stage because the formation of the products is a function of the electrode potential. Another feature of this method is that methane is converted directly to methanol. The formation of methanol may be explained in two ways. First, by the further oxidation of the methyl radical by electrolytic catalysis:



where MOH is the intermediate species in the electrochemical evolution of oxygen⁴. The second explanation involves hydrolysis of CH₃Cl:



Reaction (1) is endoergic, and reaction (3) seems to predominate. In fact, the formation of methanol was more favourable at higher pH. However, reactions (1) and (2) must occur because methanol was produced even in strongly acidic solution (pH 1). We are presently studying the mechanism in detail. The rather complicated effect of the anodizing potential on the formation of the electrolytic products may be related to the evolution of oxygen—that is, an oxygen or hydroxyl radical generated during anodization could act as a scavenger for chlorine and methyl radicals. That the most important step in this conversion process is the radical reaction in the gas phase is demonstrated by the very small amounts of the products yielded when a methane-saturated solution was used, but the electrode was completely immersed (the light then passing through the solution).

Received 12 August; accepted 7 November 1985.

1. Hatch, L. F. & Matar, S. *From Hydrocarbons to Petrochemicals* 58–61 (Gulf Publishing Company, Houston, 1981).
2. Kuhre, C. J. & Shearer, C. J. *Oil Gas J.* **85–90** (1971).
3. Nonaka, A. *Analyt. Chem.* **44**, 271–276 (1972).
4. Bockris, J. O'M. *J. chem. Phys.* **24**, 817–825 (1956).

Does partial melting reduce the creep strength of the upper mantle?

S. Karato

Ocean Research Institute, University of Tokyo, Minamidai, Nakano, Tokyo 164, Japan

The seismic low-velocity zone has been thought to correspond to a zone of partial melting that significantly reduces the creep strength, making it possible for the plates to decouple from the underlying mantle^{1,2}. This view is examined here in light of recent experimental results which show that water has more effect on the creep strength of olivine³⁻⁶ than does partial melting^{7,8} and that the former depends on the fugacity of water^{5,6}. When there is a limited amount of water, partial melting will result in depletion of water from olivine, and therefore have a hardening, rather than a softening, effect.

Recent experimental studies on olivines containing a small amount (a few per cent) of partial melt have demonstrated that it has only a minor effect on creep strength (enhancement of creep rate by a factor of $\leq 2-5$) both in water-free⁷ and water-saturated⁸ conditions. Water, however, has a more drastic effect (enhancement of creep rate by a factor of $\geq 20-50$) both on the dislocation creep³⁻⁶ and on the grain size sensitive (diffusion) creep regime⁵. The water effect has been shown to depend on the fugacity of the water (concentration of water in olivine)^{5,6}.

These observations suggest that the fugacity of water should have an important role in the rheology of the upper mantle. The fugacity of water, and, therefore, water content in olivine, depends on the amount of water present and the number of phases which can incorporate water. Silicate melts significant amounts of water⁹ so partial melting will reduce both the fugacity of water and its content in olivine if the melt fraction exceeds a certain value.

At the pressure and the temperature conditions in the low-velocity and high-conductivity zone, most hydrous phases are unstable and water can therefore exist either free and/or dissolved in melt and/or nominally anhydrous solid phase (olivine). Recent experimental work⁶ indicates that the solubility of water in olivine is $\sim 5 \times 10^{-5}$ mol per mol (or 10^{-3} wt%) at the confining pressure of 300 MPa. The solubility of water in the basaltic melts is quite large, ~ 0.5 mol per mol (or ~ 6 wt%) at the same pressure⁹. These results indicate that when the basaltic melt coexists with olivine, water will be concentrated in the melt phase with a partition coefficient $K = c_{w,m}/c_{w,s} \sim 10^4$ (at 300 MPa) where $c_{w,m}$ and $c_{w,s}$ are the solubility of water in the melt and that in the solid phase (olivine) respectively. Here we assumed that the partition of water between the melt and the olivine is independent of the fugacity of water and of the melt fraction. The actual value of K in the mantle may depend on the fugacity of water because of the possible difference in the dependence of water solubility in the melt ($c_{w,m} \sim f_{H_2O}^{1/2}$) (ref. 9) and in the olivine ($c_{w,s} \sim f_{H_2O}^m$, $m = 1/3-1/5$) (refs 5, 6) on the water fugacity. However, this possible dependence of K on water fugacity is small in the present context and will not be used in the following argument.

When the melt fraction is small, a partially-melted rock will be saturated with water. The amount of water in olivine is equal to the solubility (limit) $c_{w,s}^0$. Undersaturation of a rock with water occurs when the melt fraction exceeds a critical value $f_c = c_i/c_{w,m}^0 - K^{-1}$, where c_i is the water content of a rock and $c_{w,m}^0$ is the solubility (limit) of water in the melt. The concentration of water in olivine in this case $c_{w,s}$ is given by

$$c_{w,s}/c_{w,s}^0 = (1 + f_c(K-1))/(1 + f(K-1))$$

where f is the melt fraction, here the dependence of K on the

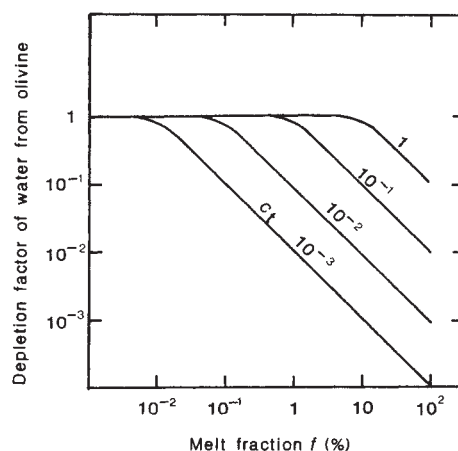


Fig. 1 Depletion factor of water in olivine due to partial melting as a function of melt fraction f for several values of total water content c_i (wt%). The water content of the source region of MORB is estimated to be $\sim 10^{-2}$ wt% based on the measured value of water content of MORB¹². Significant depletion of water and resultant hardening will occur when the melt fraction greatly exceeds the critical value, but remains less than the larger amounts ($\sim 20\%$) of melt needed to weaken the rocks¹⁸ directly.

fugacity of water is neglected. $c_{w,s}/c_{w,s}^0$ represents the ratio of the water content in olivine at an undersaturated condition to that at saturated (or melt-free) condition, and is termed a water-depletion factor. Figure 1 shows the water depletion factor as a function of melt fraction for several values of the total water content. Significant depletion of water from olivine will occur when the melt fraction exceeds a critical value. In the simplest model, where the creep rate is proportional to the concentration of water (or water-related species¹⁰), the magnitude of water depletion is proportional to the magnitude of reduction of creep rate. Therefore, partial melting will result in hardening, when the melt fraction is significantly larger than the critical value, but does not exceed the value above which the direct weakening effect of partial melting (the effect at constant water fugacity) becomes important.

The relevance of these results to the Earth depends on the water content and the melt fraction in the Earth, quantities which may vary according to the variation of the tectonic setting^{11,12}. For the source region of mid-ocean ridge basalts (MORB), the low-velocity and high-conductivity zone under the oceanic plates, the water content of the mantle (c_i) is $\sim 10^{-2}$ wt% based on the measured water content of MORB¹², assuming that MORB is formed by $\sim 20\%$ partial melting¹³. This gives $f_c = 10^{-1}\%$ a value significantly smaller than the estimated degree of partial melting in the low-velocity and high-conductivity zone¹⁴, although this estimation contains a large uncertainty. We suggest, therefore, that this hardening effect is likely to occur in the low-velocity and high-conductivity zone under the oceanic plates.

The weakening effect of partial melting observed in the laboratory is likely to depend on the grain-size sensitivity. In the Earth, the grain-size insensitive dislocation creep may also have to be considered in which the effect of partial melting will be much smaller. Therefore, the experimentally-observed weakening effect of partial melting must be regarded as the upper limit to the realistic value.

The above argument also assumes that laboratory data on the creep strength of the olivine + basalt system is applicable to the Earth. Although the melt fraction (several per cent) in these experiments^{7,8} probably represents the typical value of degree of partial melting in the Earth, the effect of other minerals must be examined before the results can be applied. Recent experiments¹⁵ on the melt-solid geometry of a partially melted peridotite indicates that the olivine-basalt-olivine boundary has the

smallest dihedral angle. This suggests that the experimental results on the olivine + basalt system give the lower limit of the creep strength of the realistic partially-melted system.

Note that so far we have considered the effect of changing the melt fraction and kept the other variables unchanged. This is certainly not realistic. An increase in the melt fraction in the Earth is nearly always associated with increase in temperature and/or in water content (or water fugacity). Both of these effects will tend to reduce the creep strength and will, therefore, partially offset the present argument. Even in these cases, the proposed mechanism will counteract to the softening effects due to the increase in temperature and/or in water content and must be regarded as a relative hardening effect.

I conclude that partial melting reduces the creep strength only in restricted conditions. The reduction in water content in olivine can be associated with partial melting in the low-velocity and high-conductivity zone under the oceanic plates, which may increase the creep strength compared with a (hypothetical) melt-free situation. Therefore, creep strengths of the low-velocity and high-conductivity zone may be significantly larger than in the case where no partial melting occurs.

This mechanism may explain the puzzling observation that the extrapolation of experimental results on water-saturated

('wet') olivine rheology predicts an embarrassingly weak creep strength of the upper mantle^{16,17}.

I thank David Kohlstedt and Prame Chopra for helpful comments, Toshitsugu Fujii for discussions on the petrological aspects, and Mitsuhiro Toriumi for stimulating discussions.

Received 2 July; accepted 25 September 1985.

1. Ringwood, A. E., in *The Earth's Crust and Upper Mantle* (ed. Hart, P. J.) 1-17 (American Geophysical Union, Washington DC, 1969).
2. Anderson, D. L. & Sammis, C. *Phys. Earth planet. Inter.* **3**, 41-50 (1970).
3. Chopra, P. N. & Paterson, M. S. *Tectonophysics* **78**, 453-473 (1981).
4. Chopra, P. N. & Paterson, M. S. *J. geophys. Res.* **89**, 7861-7876 (1984).
5. Karato, S., Paterson, M. S. & FitzGerald, J. D. *J. geophys. Res.* (in the press).
6. Mackwell, S. J., Kohlstedt, D. L. & Paterson, M. S. *J. geophys. Res.* (in the press).
7. Cooper, R. F. & Kohlstedt, D. L. *Tectonophysics* **107**, 207-233 (1984).
8. Chopra, P. N. & Kohlstedt, D. L. *EOS* **64**, 323 (1983).
9. Burnham, C. W. in *The Evolution of the Igneous Rocks* (ed. Yoder, H. S. Jr) 439-482 (Princeton University Press, 1979).
10. Hobbs, B. E. *J. geophys. Res.* **89**, 4026-4038 (1984).
11. Ito, E., Harris, D. M. & Anderson, A. T. Jr *Geochim. cosmochim. Acta* **47**, 1613-1624 (1983).
12. Moore, J. G. *Contr. Miner. Petrol.* **28**, 272-279 (1970).
13. Ringwood, A. E. *Composition and Petrology of the Earth's Mantle* (McGraw-Hill, New York, 1975).
14. Shankland, T. J., O'Connell, R. J. & Waff, H. S. *Rev. Geophys. Space Phys.* **29**, 394-406 (1981).
15. Toramaru, A. & Fujii, N. *J. geophys. Res.* (submitted).
16. Goetze, C. in *High Pressure Research, Application in Geophysics* (eds Manghnani, M. H. & Akimoto, S.) 3-23 (Academic, New York, 1977).
17. Karato, S. *Tectonophysics* **104**, 155-176 (1984).
18. van der Molen, I. & Paterson, M. S. *Contr. Miner. Petrol.* **70**, 299-318 (1979).

Partial melting zones in the crust in southern Tibet from magnetotelluric results

Van Ngoc Pham*, Danièle Boyer*, Pierre Therme*, Xue Cheng Yuan†, Li Li† & Guo Yuan Jin†

* Laboratoire de Géophysique Appliquée et Structurale (LA 235), Ecole Nationale Supérieure de Géologie, BP 40, 54501 Vandœuvre Cedex, France

† Bureau of Geophysical and Geochemical Exploration, Ministry of Geology and Mineral Resources, Xi Si, Beijing, China

Magnetotelluric studies in southern Tibet along a north-south traverse, 200 km long, crossing the Yarlung-Zangbo suture zone near Lhasa, have enabled us to construct a two-dimensional model in which layers of anomalously high conductivity exist at depths shallower than 33 km. The electrical properties of these layers are different on the two sides of the suture zone. We show here that the most plausible explanation of the high conductivity is the existence of partial melting zones in the upper part rather than the lower part of a crust ~70 km thick. This represents a powerful constraint for thermo-mechanical models of continental collisions.

The Tibetan Plateau is the largest topographical mass on the Earth, with an unusually high average elevation (5 km above sea level). Although its uplift must be related to the Himalayan continental collision and penetration of India into Asia¹, there is considerable debate over its present structure and mode of formation². The chief aim of the 3-year programme of French-Chinese cooperation for the study of Tibet and the northern Himalayas, which began in 1980, was to improve the understanding of the local geology by using new techniques and observations. The programme involved magnetotelluric (MT) exploration in southern Tibet, in the area around the Yarlung-Zangbo suture (YZS) zone which marks the boundary between the Indian continent to the south and the Lhasa Block to the north³ (Fig. 1).

MT studies provide information about the electrical properties of the crust and upper mantle, and may have an important role in characterizing contemporary tectonic activity in a continent-continent collision zone since, at different levels of the crust and upper mantle, the electrical resistivity will reflect several thermally related processes. Indeed the Tibetan Plateau is

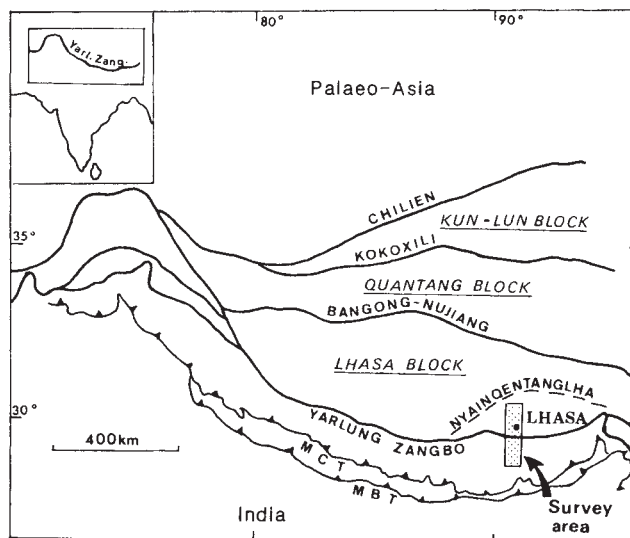


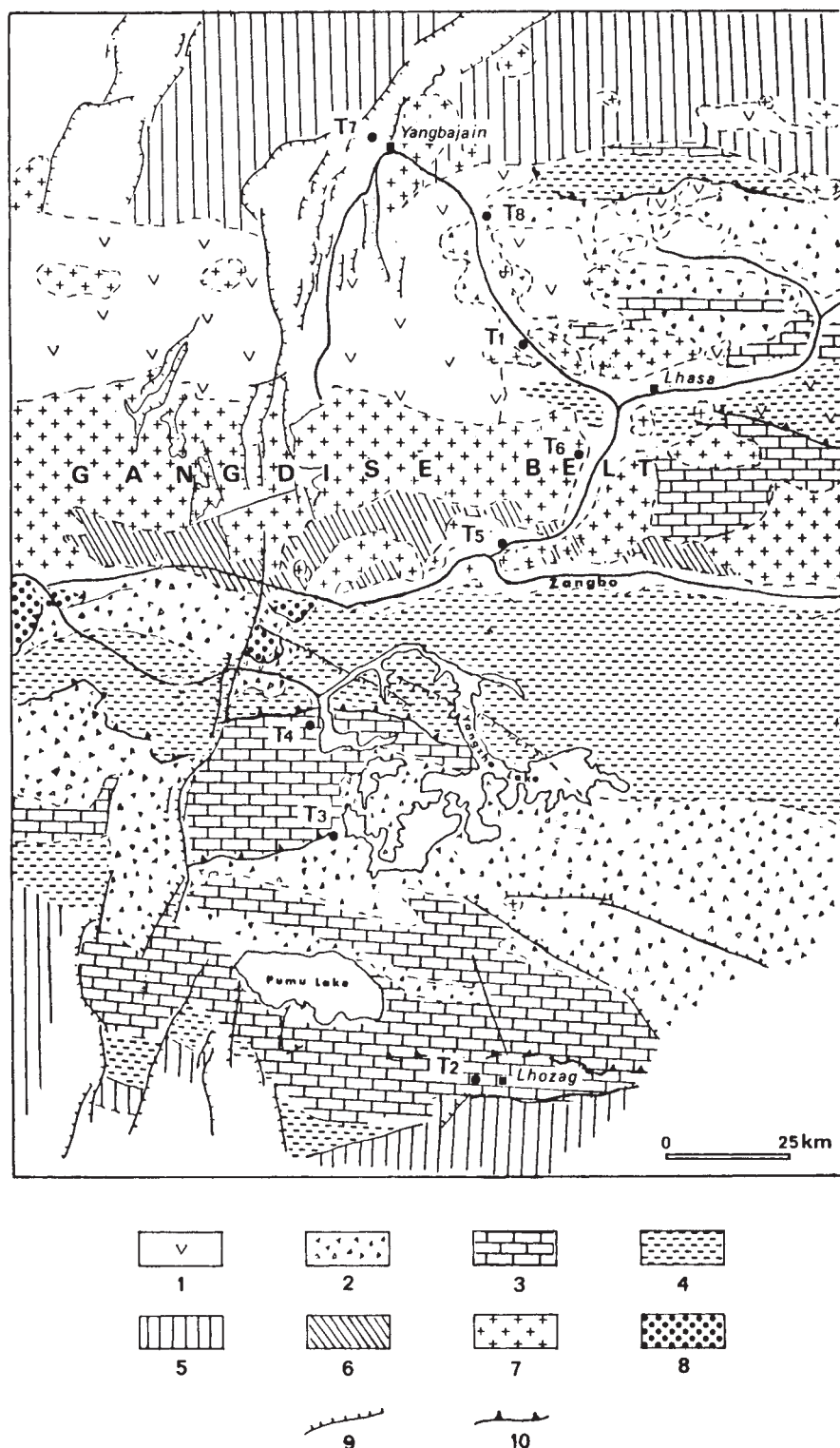
Fig. 1 Schematic tectonic map of the Tibetan Plateau and the North Indian Margin showing several independent tectonic blocks (underlined) separated by suture zones (heavy lines). Dashed line, Nyainqentangla Range, barbed lines indicate continental underthrusting within the Indian continent. MCT, Main Central Thrust; MBT, Main Boundary Thrust. Shading marks the location of the MT survey area (from ref. 3).

characterized by strong geothermal activity⁴ with more than 600 hydrothermal areas, of which 70% are concentrated in southern Tibet in the so-called 'Himalayan geothermal belt' (HGB) along both sides of the YZS zone⁵. So far, 11 hydrothermal explosions, 3 geysers and 42 boiling springs have been discovered. The temperature of the hot springs varies from 35 °C to 100 °C⁶.

Neogene calc-alkaline volcanic rocks have been found over much of Tibet⁷. Potassic lavas (tephrites, trachytes) located at Machiang, just north of the YZS⁸, have been recently dated by ³⁹Ar/⁴⁰Ar methods at 10-15 Myr (C. Coulon, personal communication), while a granite outcrop west of Kangmar area, ~80 km south of the YZS, has an age of 6-8 Myr (ref. 9).

The geothermal activity in southern Tibet is closely related to recent magmatic activity and contemporary tectonic movements. Consequently, the thermal structure of Tibet is a very important parameter for understanding the active processes in

Fig. 2 Location of MT stations on the simplified geological map of the survey area. 1, Cenozoic volcanic series; 2, Cretaceous; 3, Jurassic; 4, Trias-Liassic; 5, Palaeozoic gneiss; 6, amphibolite; 7, calc-alkaline and granite plutons; 8, ophiolites; 9, Quaternary normal fault; 10, major thrust (redrawn from ref. 14).



a continent-continent collision zone. One explanation of the geothermal activity is the existence of a partial melting zone in the crust^{4,5,6,10}, which should be detectable by MT studies because the electrical resistivity of rocks decreases as the temperature increases and falls dramatically at the melting point.

The Tibetan Plateau results from the juxtaposition of small continental blocks against Palaeo-Asia during the Mesozoic³ (Fig. 1). These are now separated from each other by ophiolitic sutures. The southernmost block, the Lhasa Block, is bounded to the north by the Bangong-Nujiang suture (BNS) and to the south by the YZS which corresponds to the last collision with the Indian continent and seems to be the most important one.

The geology (stratigraphy and tectonics) of the surrounding area of the YZS zone has been well studied during the French-Chinese joint programme (see refs 2, 3, 11-14).

The main tectonic feature of the Lhasa Block is a major strike-slip fault associated with the Nyainqentanglha range crossing near Yangbajain (Fig. 1) which divides the Lhasa Block into two parts with different tectonic histories^{15,16}. In the southern part, where half of the surveyed area is located, Cretaceous to Palaeocene volcanism, which volumetrically is more important than in northern parts is typical of an active continental margin. It reflects the northward subduction of the Indian oceanic crust beneath the Lhasa Block⁸.

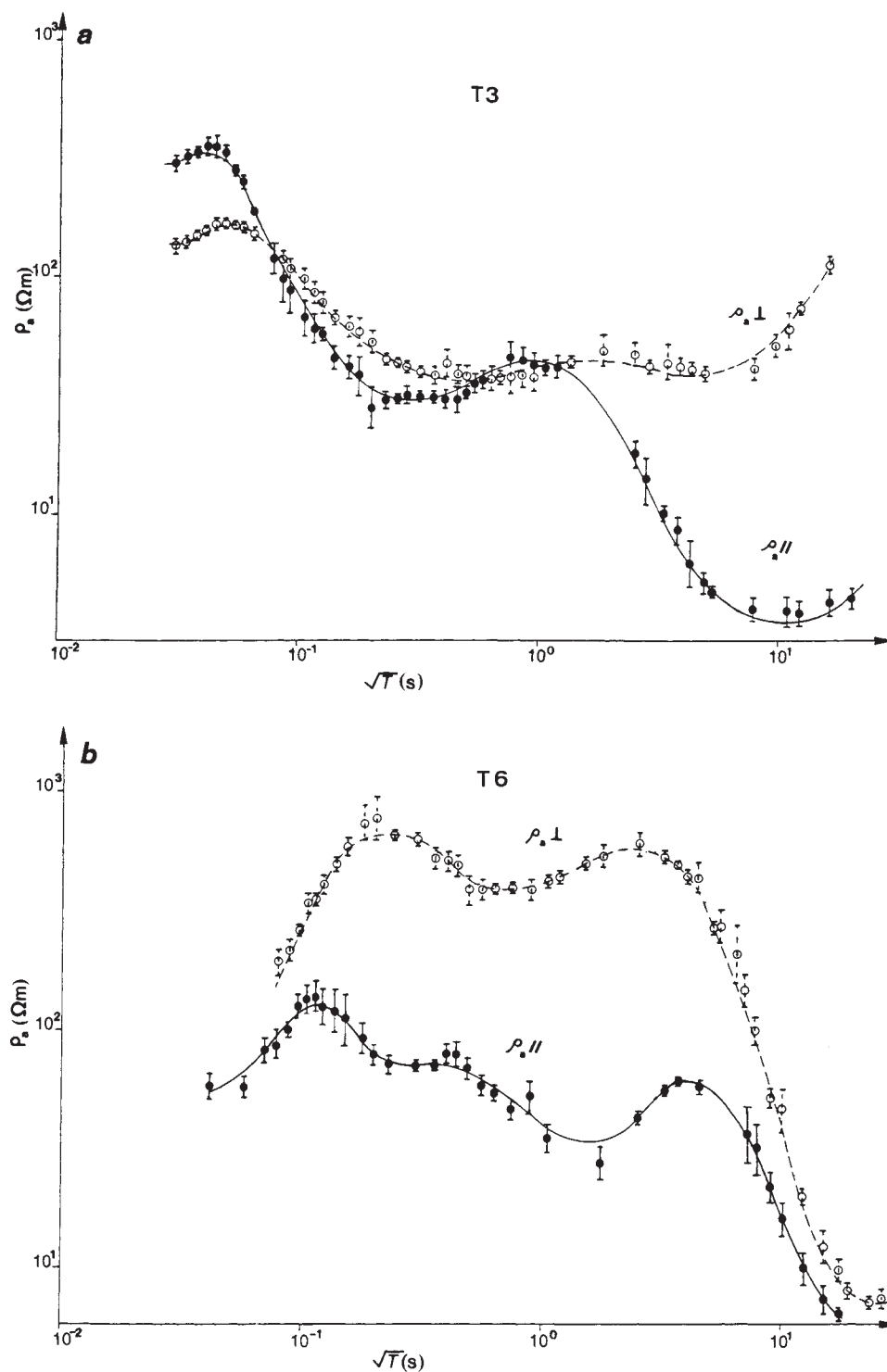


Fig. 3 MT sounding curve showing parallel and perpendicular apparent resistivities against square root of period for: *a*, station T₃; *b*, station T₆.

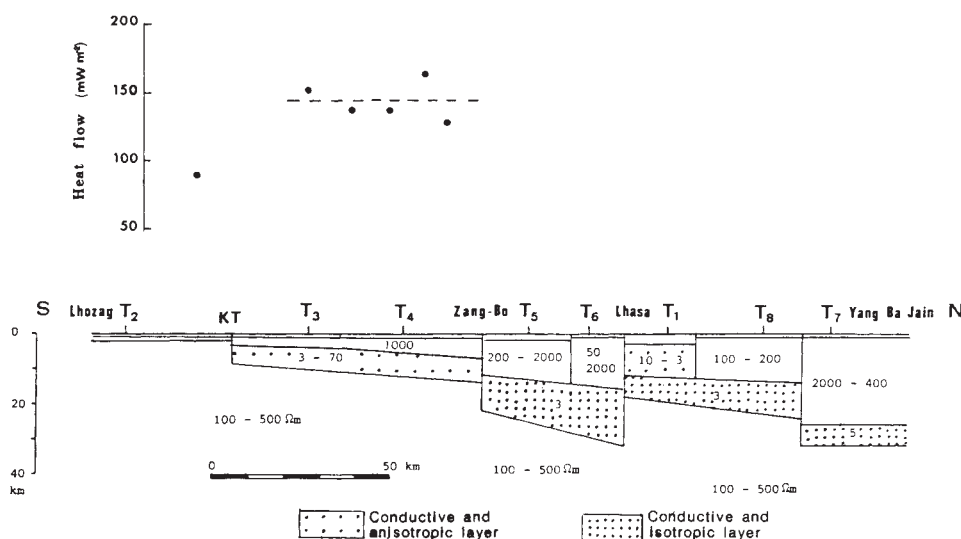
Following the mechanism proposed by Mattauer^{17,18}, the crust of the North Indian Margin has doubled in thickness (~70 km) through the formation of a crustal stacking wedge. This necessitates crust-mantle decoupling with very flat-lying thrusts. Moreover, seismic results indicate that decoupling and thrusting separate lower and upper crustal layers^{19,20}.

Along a north-south traverse over ~200 km passing near Lhasa and running on both sides of the YZS, eight deep MT soundings were carried out from Yangbajain to the north, to Lhozag to the South (Fig. 2). Three stations, T₂, T₃ and T₄, are located to the south of Zangbo in the Jurassic and Cretaceous series, and two stations, T₅ and T₆, lie must north of Zangbo in

the Gangdise granitoid belt. Farther north, T₁ is located in the Lhasa granite pluton, T₈ in the Cretaceous and Cenozoic volcanic series and T₇ in the Quaternary Yangbajain graben.

MT soundings consist of studying the variation of apparent resistivity ρ_a with the period T (or square root of period) of the natural electromagnetic field. At each station, MT signals were recorded over several ranges of periods, thereby covering more than six orders of magnitude, from 10^{-3} s to 10^3 s. Furthermore, recordings were carried out simultaneously along two perpendicular directions (east-west and north-south) to study how the electrical anisotropy was related to the micro- or macro-anisotropy and to the heterogeneity of the structure. The digital

Fig. 4 Geo-electrical section along the traverse obtained from MT two-dimensional modelling (no vertical exaggeration). In each layer, the values indicate the resistivities in Ωm . In the case of anisotropic layer, the first value corresponds to ρ'' and the second value to $\rho^\perp$. The upper part shows the heat flow values at their respective locations (after ref. 32).



data were analysed to estimate MT impedance tensor elements²¹ and apparent resistivities along two principal directions in the case of two-dimensional earth. Apparent resistivities parallel (ρ_a'') and perpendicular ($\rho_a^\perp$)²² to the tectonic strike correspond respectively to the E and H polarization cases²² in the two-dimensional modelling described below. The geological map (Fig. 2) of the area surveyed shows that the direction of the tectonic strike is near east-west and the structure is mostly two-dimensional. Table 1 summarizes information regarding the angle θ of the strike direction from the north²³ and the two-dimensionality indicator, the 'skew' index^{22,23}. These data confirm that the strike direction is near east-west at all stations except at station T₇ located in the Quaternary Yangbajain graben, striking along a north-east direction. The relatively low skew index confirms the two-dimensionality of the geological structure.

Although MT sounding curves were anisotropic at all stations, this depends on the range of periods and varies from one zone to the other along the traverse. For example, Fig. 3 shows two typical MT sounding curves corresponding to two stations, T₃ and T₆, located respectively south and north of the YZS zone.

At station T₃ (Fig. 3a), apparent resistivities are low and probably isotropic for short periods below 1 s and become strongly anisotropic for long periods above, with a minimum ρ_a'' of about 4 Ωm . This indicates the existence of a deep conductive and anisotropic zone to the south of the YZS. At station T₆ (Fig. 3b), apparent resistivities are highly anisotropic in the range of periods between 10^{-2} and 10^2 s, with ρ_a'' very low and $\rho_a^\perp$ very high. This anisotropy may be related to the micro- or macro-anisotropy provoked by the schistosity and foliation of the compressed Gangdise granitoid belt. Above a period of 100 s both ρ_a'' and $\rho_a^\perp$ decrease dramatically by several orders of magnitude, indicating the presence of a strongly conductive zone at depth.

Two-dimensional modelling of the geoelectric structure was carried out to explain the resulting MT response along the traverse, allowing for both heterogeneous and anisotropic features of the structures. The best fitting model is given in Fig. 4. A full description of the modelling procedure^{24,25} and details of these geoelectrical studies will be published elsewhere. We focus here on the origin of the highly conductive zones in the crust.

The most outstanding feature observed on the geoelectrical section (Fig. 4) is the existence of layers of very low resistivity in the upper crust, between 3 and 33 km depth. To the south of the YZS zone, a conductive and anisotropic layer is observed between 3 and 14 km depth, with parallel resistivity $\rho'' = 3 \Omega\text{m}$

and perpendicular resistivity $\rho^\perp = 70 \Omega\text{m}$. This layer dips slightly to the north and is bounded laterally by the YZS and the Kangmar thrust (KT). To the north of the YZS zone, a strangely conductive and isotropic zone is observed between 12 and 33 km depth. This zone is divided into several compartments bounded by three major discontinuities located at the central zone between two stations, bracketing each of them respectively around YZS, Lhasa and Yangbajain. Just to the north of Lhasa, and under station T₁, a more localized conductive but anisotropic zone is observed between the 3 and 13 km which overlays the conductive and isotropic zone.

Explanations for conductive zones in the Earth's crust include: (1) solid conduction through the bulk silicate material at high temperatures; (2) electrolytic conduction through fluids contained in cracks and pores; (3) conduction through a melt fraction.

Solid conduction in the bulk silicate phase due to increased temperature alone seems to be ruled out because temperatures would have to be extremely high at shallow levels in the crust, $>1,000^\circ\text{C}$ to decrease the resistivity below $10 \Omega\text{m}$ ²⁶ (ref. 26).

In considering the contribution of the electrolytic conduction, the critical factor is the porosity. Often only a few kilometres of depth are sufficient for continuing pressures to close cracks and pores. Nevertheless, even at greater depth, free water can be released from hydrated minerals during metamorphism²⁷. There are persistent fluids having water pressure almost equal to lithostatic pressure, particularly in metamorphic regions. In crustal conditions, low porosity (0.01–0.1%) may exist within fractures rather than on a grain boundary scale and would be kept open by the low effective pressure²⁸. Laboratory measurements indicate that only a little water is required to decrease

Table 1 Characteristics of the two-dimensionality of the structures in the range of periods 1–1,000 s

Station	θ^* (deg)	Skew†
T ₁	$+100 \pm 4$	0.4 ± 0.2
T ₂	$+90 \pm 5$	0.2 ± 0.2
T ₃	$+102 \pm 5$	0.4 ± 0.2
T ₄	$+98 \pm 4$	0.4 ± 0.1
T ₅	$+91 \pm 4$	0.3 ± 0.2
T ₆	$+100 \pm 3$	0.4 ± 0.2
T ₇	$+40 \pm 5$	0.4 ± 0.1
T ₈	$+78 \pm 4$	0.4 ± 0.2

* Mean value of the angle of structural strike clockwise from the north and r.m.s. error.

† Mean value of the skew index and r.m.s. error.

resistivity by many orders of magnitude²⁹. Consequently, low resistivity values could be produced by such environments.

However, in active tectonic regions characterized by strong geothermal activity, the main contribution to lowered resistivities may be the partial melting of rock components, especially if the conductive zone has a very low resistivity, below 10 Ω m. This seems to apply to the area surveyed in southern Tibet³⁰.

The electrical characteristics of the conductive zones are quite different in the southern and northern parts of the YZS zone. To the south, the conductive zone is located at a relatively shallow depth (3–14 km) under stations T₃ and T₄ and is anisotropic. Although the presence of water-rich fluid with a strong content of conducting solutes could explain its high conductivity, because of the very low value of the parallel resistivity ($\rho'' = 3 \Omega$ m), there must be another contributory factor. Recent heat flow measurements carried out in the same area show that heat flow rises sharply from 91 to 146 mW m⁻² precisely above the southern boundary of the conductive zone (between stations T₂ and T₃), and remains approximately constant over a distance >40 km towards the YZS zone³¹ (Fig. 4). This suggests that the heat source has large horizontal dimensions. A model using both the amplitude and the spatial characteristics of the thermal anomaly shows that the heat source must be younger than 1 Myr and shallower than 10 km (ref. 32). This agrees well with the spatial characteristics of the conductive zone outlined by MT results. This anomalous source of heat could be a hot intrusive body³¹; however, the anisotropic character of the conductive zone makes it more likely that the interstitial melt was acting in series with host rocks, intruding along fractures parallel to the tectonic strike in an east–west direction, and producing the 'macro-anisotropy' observed from MT results in this area.

To the north of the YZS zone, the most plausible explanation for the strongly conductive and isotropic zones between 12 and 33 km depth is the existence of a large melt fraction concentrated along intergranular boundaries. With the onset of melting there is a dramatic increase in the electrical conductivity³³. The degree to which partial melt increases the total conductivity of a material depends on the geometry of the melt distribution, particularly on the connectivity of the fluid phase throughout the solid matrix^{34,35}, and on the presence of water^{29,33}. In the presence of excess water, granite melts in the crust³⁶ at a temperature between 600 and 650 °C. Either these temperatures exist under the northern part of the YZS zone at the levels of the conductive zones, where a large amount of free water was released by different processes of metamorphism, or the primary melt was formed at greater depth and migrated vertically over a large distance. In any case, a resistivity as low as 3–5 Ω m and the isotropic character of these conductive zones imply the existence of a large melt fraction. Another interesting observation on the geoelectrical section of Fig. 4 is the existence of a conductive and anisotropic zone overlying the conductive and isotropic zone under station T₁. Again, one must invoke the interstitial melt which produces the observed macro-anisotropy. But here, the perpendicular resistivity is slightly lower than the parallel resistivity, contrary to what occurs in the southern part of the YZS zone. This feature can be explained by the intrusion of melt along recent north–south normal faults observed around Lhasa^{37,38} resulting from a dominant tectonic process of east–west extension during the Quaternary^{2,11}.

The MT results reported here suggest that regional melting conditions are reached at relatively shallow crustal depths in southern Tibet, contrary to previous reports; in particular, the study of long-period seismic wave propagation^{10,39} suggested a possible partial melting zone in the lowermost part of the crust, that is >40–50 km depth in the case of a 70-km thick crust. The actual MT results show that partial melting exists only in the upper crust. No conductivity anomaly has been detected in the lower crust or the upper mantle. The excellent agreement with

heat flow measurements in the southern part of the YZS zone strengthens our hypothesis for the origin of conductive zones.

The existence of partial melting zones at relatively shallow depth is evidently related to the recent magmatic and geothermal activities in southern Tibet. Its localization in the upper rather than the lower crust represents a powerful constraint for thermo-mechanical models of continental collisions.

This work is sponsored by INAG and CNRS (France) and the Ministry of Geology and the Academia Sinica of China. We thank all Chinese colleagues of the Ministry of Geology who participated in the fieldwork during two missions in Tibet.

Received 18 July; accepted 28 October 1985.

1. Molnar, P. & Tapponnier, P. *Science* **189**, 419–426 (1975).
2. Tapponnier, P. *et al. Nature* **294**, 405–410 (1981).
3. Allègre, C. J. *et al. Nature* **307**, 17–22 (1984).
4. Wang, J. Y. *et al. J. Volcanol. geotherm. Res.* **9**, 57–76 (1981).
5. Tong W. & Zhang, M. T. *Proc. Symp. Qinghai-Xizang (Tibet) Plateau*, 841–846 (Science Press Beijing, 1981).
6. Wei, S. Y. *et al. Proc. Symp. Qinghai-Xizang (Tibet) Plateau*, 865–874 (Science Press Beijing, 1981).
7. Kidd, W. S. F. *EOS* **56**, 453 (1975).
8. Coulon, C., Maluski, H. & Wang, S. C. *Int. Symp. Himalayan Geol. Abstr.*, Cheng Du (Chinese Academy for Geological Sciences Beijing, 1984).
9. Maluski, H. *Int. Symp. Himalayan Geol. Abstr.*, Cheng Du (1984).
10. Maluski, H. *Int. Symp. Himalayan Geol. Abstr.*, Cheng Du (1984).
11. Toksöz, M. N., Buck, W. R. & Hsu, A. T. *Proc. Symp. Qinghai-Xizang (Tibet) Plateau*, 847–857 (Science Press Beijing, 1981).
12. Tapponnier, P., Mercier, J. L., Armijo, R., Tonghin, H. & Ji, Z. *Nature* **294**, 410–414 (1981).
13. Nicolas, A. *et al. Nature* **294**, 414–417 (1981).
14. Burg, J. P. & Chen, G. M. *Nature* **311**, 219–223 (1984).
15. Burg, J. P. thesis, Univ. Montpellier (1983).
16. Achache, J., Courtillot, V. & Zhou, Y. X. *J. geophys. Res.* **89**, 10311–10339 (1984).
17. Tapponnier, P., Peltzer, G. & Armijo, R. in *Collision Tectonics* (eds Ries, A. & Coward, M.) (Geological Society, London, in the press).
18. Mattauer, M. *Collision Tectonics* (Spec. Pap. Geological Society London, 1985).
19. Mattauer, M. C. *r. hebdom. Séanc. Acad. Sci., Paris* **296**, 481–486 (1983).
20. Hirn, A. *et al. Nature* **307**, 23–25 (1984).
21. Hirn, A. *et al. Nature* **307**, 25–27 (1984).
22. Sims, W. E., Bostick, F. X. & Smith, H. W. *Geophysics* **36**, 938–942 (1971).
23. Vozoff, K. *Geophysics* **37**, 98–141 (1972).
24. Swift, C. M. Jr thesis, Massachusetts Inst Technol. (1967).
25. Doucet, D. & Pham, V. N. *Geophys. Prosp.* **32**, 292–316 (1984).
26. Chung, S. G. thesis, INPL, Nancy (1984).
27. Stanley, W. D. *et al. J. geophys. Res.* **82**, 2501–2514 (1977).
28. Berdichevsky, M. N. *et al. Beitr. Geophys.* **81**, 187–196 (1972).
29. Shanland, T. J. & Ander, M. E. *J. geophys. Res.* **88**, 9475–9484 (1983).
30. Olhoeft, G. R. *J. geophys. Res.* **86**, 931–936 (1981).
31. Pham, V. N., Boyer, D. & Therme, P. *Terra Cognita* **3**, 270 (1983).
32. Francheteau, J. *et al. Nature* **307**, 32–36 (1984).
33. Jaupart, C., Francheteau, J. & Shen, X. J. *Geophys. J. R. astr. Soc.* **81**, 131–155 (1985).
34. Lebedev, E. B. & Khitarov, N. I. *Geokhimiya* **3**, 195–201 (1964).
35. Waff, H. S. *J. geophys. Res.* **79**, 4003–4010 (1974).
36. Hermance, J. F. *Geophys. Res. Lett.* **6**, 613–616 (1979).
37. Wyllie, P. J. *Geophys. Monogr.* **14**, 279–301 (1971).
38. Chen, G. M. & Li, T. D. *Mission franco-chinoise 1980*, 357–383 (CNRS, France, 1984).
39. Armijo, R. *et al. EOS* **63**, 1093 (1982).
40. Bird, P. & Toksöz, M. N. *Nature* **266**, 161–163 (1977).

Isoleucine stereoisomers on the Earth

Jeffrey L. Bada*, Meixun Zhao*, Spencer Steinberg† & Edward Ruth†

* Amino Acid Dating Laboratory (A-012B), Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA

† Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024, USA

The amino acid L-isoleucine (L-Ile) undergoes epimerization at the α -carbon producing D-allo-isoleucine (D-allo-Ile), an amino acid not generally found in living organisms. In fossils and oceanic sediments, this reaction has a half-life, that is, the time required to reach D-allo-Ile/L-Ile ~ 0.4 , of 10^5 – 10^7 yr (ref. 1). Two other possible isoleucine stereoisomers, D-isoleucine (D-Ile) and L-allo-isoleucine (L-allo-Ile), have not been previously detected in fossils or sediments. These amino acids could be produced by epimerization of L-Ile and D-allo-Ile at the β -carbon². However, the β -carbon is insulated from the amino and carboxy groups which by electron-withdrawing effects and resonance stabilization promote epimeriza-

ation at the α -position³. Thus, the β -epimerization rates of the isoleucine stereoisomers are predicted to be extremely slow in comparison to their α -epimerization rates. However, no kinetic studies have been conducted. To assess whether the D-Ile and L-Ile could be produced on the Earth from β -epimerization of L-Ile and D-allo-Ile, we have measured the β -epimerization rate at 250 °C in aqueous solution, and determined the isoleucine stereoisomeric composition in fossil molluscs of various geological ages.

The four isoleucine stereoisomers are shown in Fig. 1 as are their interconversion pathways by epimerization at the α - and β -carbons. The α -epimerization rates of D-Ile and L-allo-Ile would be expected to be similar to those of the diastereomeric pair L-Ile and D-allo-Ile. Interconversion of the enantiomers L-Ile and D-Ile, and D-allo-Ile and L-allo-Ile, should not occur as this would require simultaneous inversion at two chiral centres.

Aliquots of L-Ile in phosphate buffer (pH 7.4 at 25 °C) were sealed in Teflon-lined steel reaction vessels and heated at 250 °C for periods ranging from 5 h to 3 days. The pH of the solution after heating was identical to the initial pH. Based on the extrapolation of the pH versus temperature equations⁴ for phosphate buffers determined from measurements between 0 and 90 °C, we estimate that the pH of our solution at 250 °C was ~ 8.4 . The heated solutions were desalted using cation exchange chromatography and then analysed using the TPC (trifluoroacetyl-L-prolyl chloride) gas-chromatography method⁵. The results of the 250 °C experiment are shown in Fig. 2. The estimated half-life at 250 °C and pH ~ 8.4 for the β -epimerization of L-Ile is ~ 11 days, whereas that for D-allo-Ile is ~ 25 days. The isoleucine decomposition rate was found greatly to exceed the rate of β -epimerization. In the sample heated for 70 h, only 2% of the initial isoleucine remained. Epimerization at the α -carbon is rapid at 250 °C. After only 5 h, D-allo-Ile/L-Ile = 1.3, the approximate equilibrium ratio of these two diastereomers³. The half-life at neutral pH for α -epimerization of L-Ile estimated by extrapolation of measurements⁶⁻⁸ made between 90 and 180 °C is $\sim 3-6$ min at 250 °C. The aqueous solution measurements indicate that the ratio of the α - to β -epimerization rates of isoleucine is 10^3-10^4 in the neutral pH region at 250 °C. If this ratio is also applicable to geochemical systems, then the isoleucine β -epimerization rate should have a half-life of $\sim 10^8-10^{11}$ yr on the Earth's surface. This half-life implies that little if any β -epimerization of L-Ile and D-allo-Ile would occur even over the entire age of the Earth. Because the isoleucine decomposition rate is greater than β -epimerization, isoleucine would also decompose faster than D-Ile and L-allo-Ile would be produced through β -epimerization. The slow isoleucine β -epimerization rate and the rare occurrence of D-Ile and L-allo-Ile in the biota^{9,10} suggest that L-Ile and D-allo-Ile should be the only isoleucine stereoisomers in the Earth's amino-acid reservoir. In contrast, all four isoleucine stereoisomers are present in approximately equal amounts in amino-acid rich meteorites¹¹, which is expected since the meteoritic amino acids are produced by abiotic processes¹².

For a better evaluation of the isoleucine geochemical β -epimerization rate we analysed fossil molluscs with ages of $\sim 50,000$ and ~ 1 Myr (Table 1). Both D-Ile and L-allo-Ile were detected in the oldest sample and these isoleucine stereoisomers seem to increase in abundance with increasing geological age. The β -epimerization rate of L-Ile is $\sim 2-3$ times that of D-allo-Ile, which is consistent with the rates determined in our 250 °C experiment. Using the data in Table 1, we calculate that the ratio of the α - to β -epimerization rates in these calcareous fossils is $\sim 10-50$. The rate of isoleucine epimerization at the β -carbon in the fossil mollusks therefore appears to be substantially enhanced compared with the aqueous solution measurements at 250 °C and neutral pH.

The apparent enhanced rate of isoleucine β -epimerization in the carbonate fossils may be due to several factors. The pH dependence of the α/β -epimerization rate ratio has not been

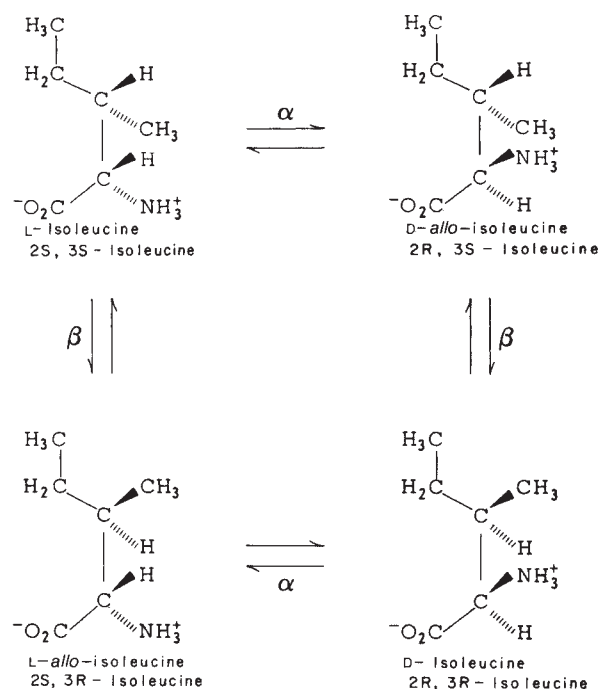


Fig. 1 Structures of the various isoleucine stereoisomers¹⁸. Both the common and R-S system names are given. Interconversion pathways at the chiral centres are indicated by α and β .

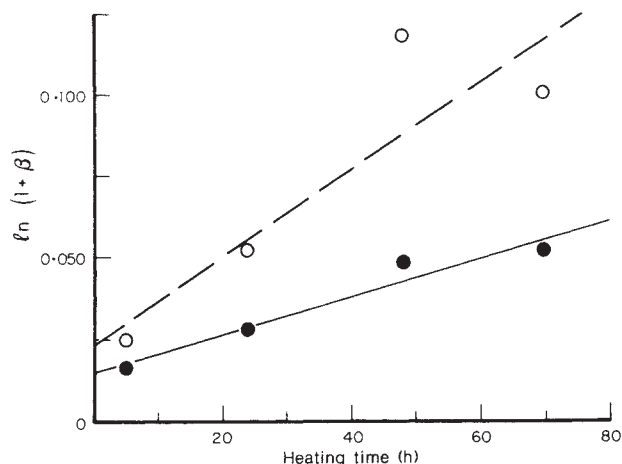


Fig. 2 An irreversible first-order kinetic plot³ for β -epimerization of L-Ile and D-allo-Ile at 250 °C and pH ~ 8.4 . Although the starting solutions contained only L-Ile, D-allo-Ile was rapidly produced through α -epimerization of L-Ile; the estimated half-life for α -epimerization at 250 °C, pH 7.6 is $\sim 3-6$ min (see text). The β term in the log function equals either D-Ile/D-allo-Ile (●) or L-allo-Ile/L-Ile (O). The slopes of the lines equal k_β , the respective β -epimerization rates of L-Ile and D-allo-Ile. The half-life for β -epimerization $\sim 1/\ln 2 \cdot 2/k_\beta$. The stereoisomeric ratios were determined by synthesizing the TPC amino-acid methyl esters⁶. Separations were performed using the gas chromatographic system and conditions described elsewhere¹⁹. All four isoleucine stereoisomers are baseline separated by this analytical method. Stereoisomeric ratios were calculated from peak heights. The detailed analytical procedures and representative gas chromatographic results will be published elsewhere.

Table 1 Isoleucine stereoisomers in fossil molluscs

Sample	Approximate age (yr)	Total <i>a</i> Ile/Ile*	D-Ile/D- <i>a</i> Ile†	L- <i>a</i> Ile/L-Ile‡
<i>Saxidomus</i> ‡ Cape Blanco, Oregon	50,000	0.19 (0.18)	0.00§	0.00§
<i>Mercenaria</i> ‡ Calabash, North Carolina	Early Pleistocene 10 ⁶	1.12 (1.10)	0.06	0.16

* The total ratio, (D-*a*Ile + L-*a*Ile)/(D-Ile + L-Ile) was determined on a Beckman 119 amino-acid analyser. The number in parentheses is the average ratio determined by ion-exchange chromatography (see ref. 20).

† Computerized gas chromatography/mass spectrometry with specific ion monitoring and peak-area integration was used to determine the stereoisomeric ratios. Gas chromatography methodology is the same as that used in the 250 °C experiment shown in Fig. 2. We estimate that these ratios have an uncertainty of $\sim \pm 10\%$.

‡ Specimens used in the interlaboratory comparison study of Wehmiller²⁰. We also analysed the third sample in the Wehmiller study, such as, the 100,000–250,000-yr old *Mercenaria* (clam) shell from Charleston, South Carolina. Although we were able to detect trace quantities of D-Ile and L-*a*Ile in this sample by gas chromatography using flame ionization detection, we were unable to confirm their presence by specific ion monitoring because of the small quantities.

§ Neither D-Ile or L-*a*Ile were detectable in this sample.

determined. The effective pH of the carbonate matrix could result in a α/β epimerization rate ratio which is less than we measured in our 250 °C, pH ~ 8.4 experiment. In addition, because the temperature dependence of β -epimerization is unknown, the α/β rate ratio could be substantially different at environmental temperatures compared with that at 250 °C. Another factor could be catalysis by mineral surfaces. Catalysts are apparently important in epimerization reactions at hydrophobic chiral centres^{13,14}, and thus carbonate may selectively catalyse epimerization at the β -position. Finally, D-Ile and L-*a*Ile in fossils might be produced from β -methyl- α -ketovaleric acid, the α -keto acid analogue of isoleucine, rather than through β -epimerization. This α -keto acid undergoes rapid enolization at slightly basic pH values, resulting in racemization at the β position¹⁵. If L- β -methyl- α -ketovaleric acid was present initially in small quantities in the carbonate matrix, its racemization and subsequent transamination could result in the production of the D-Ile and L-*a*Ile over geological time. However, β -methyl- α -ketovaleric acid has not been reported in carbonate fossils, although several simpler α -keto acids have been detected¹⁶.

Our fossil data suggest that all four isoleucine stereoisomers are present in ancient biogenic carbonates. We estimate that the half-life for β -epimerization in the calcareous matrix is ~ 5 –10 Myr in temperate climatic environments. In ocean sediments, the half-lives could be of the order of 100 Myr assuming that the Arrhenius activation energy E_a for β -epimerization is similar to that at the α -position^{6,8}, for example, $E_a \approx 30$ kcal mol⁻¹. The β -epimerization reaction of isoleucine has important geochemical implications. It could provide an index of the degree of preservation of isoleucine in ancient fossils and sediments. For example, syngenetic isoleucine in Precambrian carbonaceous sedimentary rocks should exist as an equilibrium mixture containing all four isoleucine stereoisomers. The β -epimerization of isoleucine could also be useful in dating fossils and sediments in which the other amino acids have been completely racemized. The geological interval where β -epimerization of isoleucine could be used as a geochronological method would greatly exceed that generally applicable to amino acid racemization dating which at present is limited to Pleistocene, and perhaps Pliocene, samples^{1,17}.

We thank the donors of the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research. Research at UCLA was supported by a grant from the Department of Energy (award DE-AT03-76EV70134).

Received 23 July; accepted 17 October 1985.

1. Bada, J. L. *A. Rev. Earth planet. Sci.* **13**, 241–268 (1985).
2. Kvenvolden, K. A. *A. Rev. Earth planet. Sci.* **3**, 183–212 (1975).
3. Bada, J. L. in *Chemistry and Biochemistry of the Amino Acids* (ed. Barrett, G. C.) 399–414 (Chapman and Hall, London, 1985).
4. Bates, R. G. *Determination of pH* 76 (Wiley, New York, 1964).
5. Hoopes, E. H., Peltzer, E. T. & Bada, J. L. *J. chromatogr. Sci.* **16**, 556–560 (1978).

6. Bada, J. L. *Adv. Chem. Ser. No.* **106**, 309–331 (1971).
7. Dungworth, G., Vincken, N. J. & Schwartz, A. W. *Adv. org. Geochem.* **1973**, 689 (1974).
8. Smith, G. G., Williams, K. M. & Wonnacott, D. M. *J. org. Chem.* **43**, 1–5 (1978).
9. Bodansky, M. & Perlman, D. *Nature* **218**, 291–292 (1968).
10. Yajima, T., Grigg, M. A. & Katz, E. *Archs biochem. Biophys.* **151**, 565–575 (1972).
11. Cronin, J. R., Gandy, W. E. & Pizzarello, S. *J. molec. Evol.* **17**, 265–272 (1981).
12. Wolman, Y., Haverland, W. J. & Miller, S. L. *Proc. natn. Acad. Sci. U.S.A.* **69**, 809–811 (1972).
13. Enslinger, A. et al. *Adv. org. Geochem.* **1973**, 245–260 (1974).
14. Sieskind, O., Joly, G. & Albrecht, P. *Geochem. cosmochem. Acta* **43**, 1675–1679 (1979).
15. Meister, A. *J. biol. Chem.* **190**, 269–276 (1951).
16. Steinberg, S. M. & Bada, J. L. *Geochem. cosmochem. Acta* **47**, 1481–1486 (1983).
17. Wehmiller, J. F. in *Quaternary Dating Methods* (ed. Mahaney, W. C.) 171–193 (Elsevier, Amsterdam, 1984).
18. Cohn, W. E. *Meth. Enzymol.* **106**, 3–17 (1984).
19. Boehm, M. F. & Bada, J. L. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5263–5266 (1984).
20. Wehmiller, J. F. *Quat. Res.* **22**, 109–120 (1984).

Bicadinane, a C₃₀ pentacyclic isoprenoid hydrocarbon found in crude oil

H. C. Cox^{*||}, J. W. de Leeuw^{*}, P. A. Schenck^{*},
H. van Koningsveld^{*}, J. C. Jansen^{*}, B. van de Graaf^{*},
V. J. van Geerestein[†], J. A. Kanter[‡], C. Kruk[‡]
& A. W. H. Jans[§]

^{*} Department of Chemistry and Chemical Engineering, Delft University of Technology, Delft, The Netherlands

[†] Laboratory of Crystal and Structural Chemistry, Utrecht State University, Utrecht, The Netherlands

[‡] Laboratory of Organic Chemistry, University of Amsterdam, Amsterdam, The Netherlands

[§] Free University, Brussels, Belgium

It is generally accepted that saturated polycyclic hydrocarbons such as steranes and triterpanes, present in crude oils and sediments, originate from naturally occurring steroids and triterpenoids. The occurrence of individual steranes and triterpanes can be used to relate the oil or sediment sample to the biological material from which it is derived. All previously known saturated C₃₀ pentacyclic hydrocarbons could be related to a naturally occurring triterpenoid. Here we report that the structure of a novel C₃₀H₅₂ pentacyclic isoprenoid hydrocarbon present in some Far Eastern crude oils has been established as 2,6a,12-trimethyl-4,9-diisopropyl-perhydrobenzo[de]naphthacene by joint application of chromatographic methods, mass spectrometry (MS), nuclear magnetic resonance (NMR), molecular mechanics and X-ray diffraction. As far as we know, this is the first report of the occurrence in nature of a pentacyclic C₃₀ isoprenoid compound with a perhydrobenzo[de]naphthacene skeleton. It can be considered as a dimer of the sesquiterpane cadinane. A biosynthesis of the skeleton by direct cyclization of squalene seems impossible in this case.

^{||} To whom correspondence should be addressed at: Organic Geochemistry Unit, De Vries van Heystplantsoen 2, 2628 RZ Delft, The Netherlands.

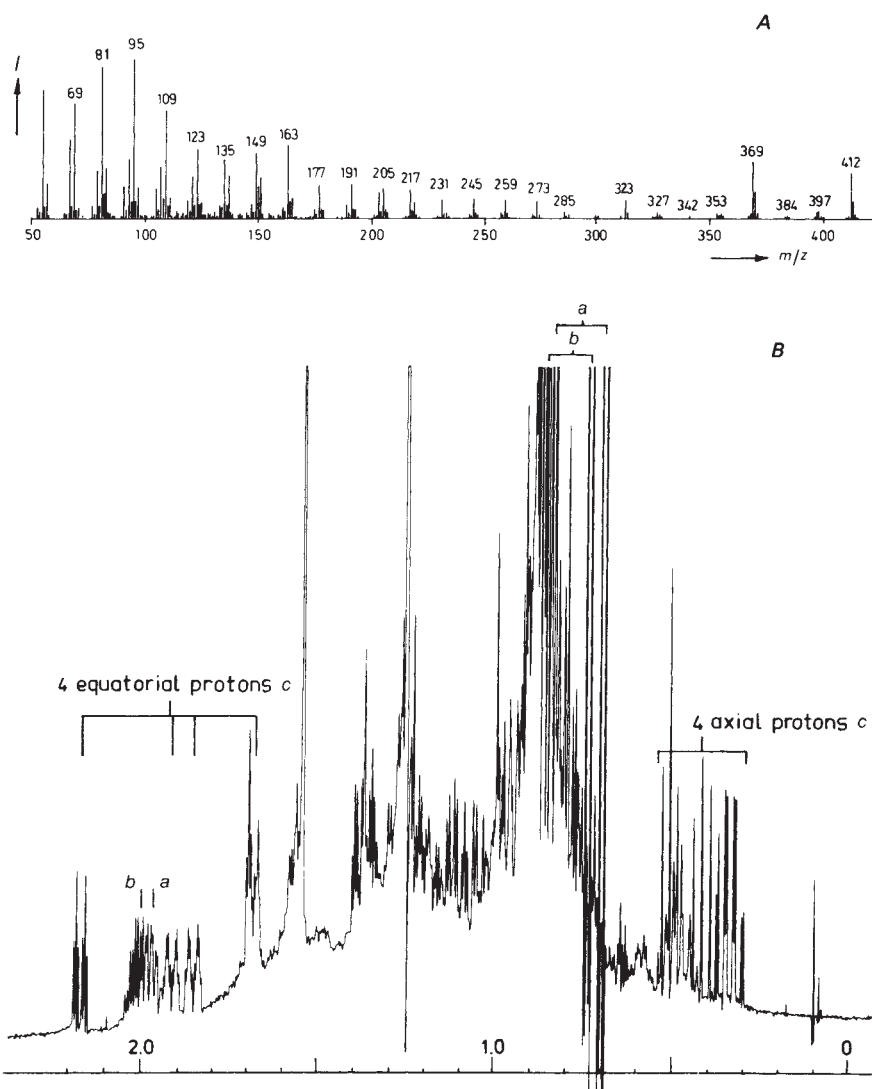


Fig. 1 A, Mass spectrum of compound T from crude oil from well W. Prabumulih-3; B, 500 MHz ¹H-NMR spectrum of compound T. Isopropyl group a, $\delta\text{CH}_3 = 0.68$ and 0.83 , $\delta_{\text{tert H}} = 1.96$, $J = 6.9$ Hz. Isopropyl group b, $\delta\text{CH}_3 = 0.72$ and 0.84 , $\delta_{\text{tert H}} = 2.10$, $J = 6.9$ Hz. Four high-field protons c, $\delta = 0.3\text{--}0.5$, each coupled to a geminal equatorial proton in the region $\delta 1.7\text{--}2.2$.

In 1982, Grantham *et al.*¹ reported the occurrence of three new compounds relatively abundantly present in the saturated hydrocarbon fractions of several Far Eastern crude oils. Based on gas chromatographic/low resolution mass spectrometric (GC/LRMS) data, they preliminarily characterized these compounds (labelled W, T and R) as C_{30} pentacyclic hydrocarbons, presumably originating from Tertiary land-plant resins. Because of their selective occurrence in these oils, these components are thought to be important biological markers with a potential use in oil/oil and oil/source rock correlation studies. Therefore, we have undertaken the structural identification of one of these compounds, T.

Very small amounts of spontaneously formed crystals, present in molecular distillation fractions of saturated hydrocarbons isolated from a South Sumatra crude oil, were analysed by GC and GC/LRMS. GC-retention time and mass spectral data of this crystalline compound were the same as those obtained for compound T present in a crude oil from the W. Prabumulih-3 well. The crystals were carefully rinsed with *n*-hexane and a single crystal was subjected to X-ray diffraction analysis using an Enraf-Nonius CAD-4 single-crystal diffractometer. For 7,114 independent reflections the intensities were collected using graphite-monochromatized $\text{MoK}\alpha$ radiation. The space group was established as $\text{P}2_12_12_1$ with $a = 10.106(1)$, $b = 17.831(3)$, $c = 29.666(5)$ Å, with two independent molecules in the asymmetric unit. An additional translational pseudosymmetry seriously hampered a straightforward determination of the structure by conventional direct methods. To overcome these difficul-

ties, more structural information was necessary to limit the number of search models for the application of a recently developed Patterson search program². Therefore, larger amounts of pure compound T were required. Following the procedure for the isolation of triterpanes previously described by Hills and Whitehead³, 4 mg of almost pure compound T was isolated from 100 cm³ of crude oil from the well W. Prabumulih-3. Compound T was obtained from the appropriate molecular distillation fraction (checked by GC analysis) by column chromatography on SiO_2 and Al_2O_3 , treatment with 5-Å molecular sieves, repeated column chromatography on Al_2O_3 with *n*-hexane as an eluent and urea clathration. Resulting crystalline material was purified by recrystallization from *n*-heptane. Some 2.5 mg of compound T, isolated and purified in this way, was used for NMR analyses. Another aliquot was investigated by MS. The mass spectrum is shown in Fig. 1A. Exact mass measurements of the molecular ion at 412 daltons yielded a value of 412.405, which corresponds to a calculated value of 412.406 for $\text{C}_{30}\text{H}_{52}$. A linked scan with B^2/E constant, performed for $m/z = 313$, showed that the principal mother ions of this fragment are $m/z = 369$ ($\text{M}^+ - 43$) and $m/z = 412$ (M^+). The mass spectral data suggest the presence of at least one isopropyl group. The loss of 99 daltons as indicated by the relative abundance of $m/z = 313$ probably reflects a two-stage fragmentation process, thus making the presence of a C_7H_{15} side chain less likely (compare with ref. 1).

Figure 1B shows the 500-MHz ¹H-NMR spectrum. Double-resonance experiments and ¹H shift-correlated two-dimensional



Fig. 2 A, Cadinane; B, eudesmane.

NMR (COSY)^{4,5} showed the presence of two non-equivalent isopropyl groups and four CH₂ groups in the 2-position of a 1,3-di-equatorially substituted cyclohexane ring⁶. Analysis of 62.8-MHz proton-decoupled ¹³C-NMR spectra indicated that all 30 carbons have sp³ hybridization. Information about the multiplicity of the carbons could be obtained by using the APT⁷ and DEPT⁸ pulse sequences. A combination of these experiments with a 125-MHz ¹³C-¹H shift-correlated two-dimensional NMR spectrum⁹ revealed the presence of 7 CH₃-, 9 CH₂-, 13 CH-groups and 1 quaternary carbon atom. The NMR techniques used are described in detail elsewhere^{4,5,7-9}.

From these NMR data it became clear that the C₃₀H₅₂ compound T is not a triterpane in the usual sense because only one quaternary carbon atom is present in the structure. The elemental formula and the presence of two isopropyl groups and three additional methyl groups indicate that the skeleton is pentacyclic and consists of 21 carbon atoms. A literature search for natural compounds with structural features as revealed by the NMR and MS data (high-field protons, isopropyl substituents, only one quaternary carbon atom) led us to consider saturated sesquiterpenoid hydrocarbons, especially cadinane and eudesmane¹⁰ (Fig. 2). These compounds, although containing only 15 carbon atoms, possess an isopropyl group, one or no quaternary carbon atom and the characteristic CH₂ group with a high-field proton. Moreover, the mass spectra of these compounds are very similar to the lower part of the mass spectrum of compound T. Further circumstantial evidence for the presence of these sesquiterpenoid moieties in compound T was obtained from the relative abundance of cadinane in the crude oil as revealed by GC/MS analysis. Additional information resulted from Curie point pyrolysis/GC/MS¹¹ of the Lumapas resin, shown to be a source of components R, T and W on heating¹. Many sesquiterpenoid hydrocarbons with at least one isopropyl group were encountered in the pyrolysate. These additional data suggested that compound T could be considered as a 'dimer' of cadinane or eudesmane. Coupling of two decalines through a methene group structured in a cyclopentane or cyclohexane ring indeed yields a pentacyclic skeleton of 21 carbon atoms. A pentacyclic structure with five six-membered rings was preferred, as the 125-MHz proton-coupled ¹³C-spectrum did not reveal ¹J_{C-H} values greater than 126 Hz. The Patterson map calculated from the X-ray data precluded structures with a central cyclopentane ring. Only six skeletons containing 21 carbon atoms and consisting of five six-membered rings exist (Fig. 3A-F). Structures D, E and F in Fig. 3 are less likely because they cannot be constructed from two decaline systems. Molecular mechanical calculations using the MM2 force field¹² revealed steric energies for the all-*trans* structures A, B and C of 26, 28 and 36 kcal mol⁻¹ respectively. Therefore, only structures A and B were considered to be promising search models. After Patterson searches with the Patsee program, skeleton B appeared to be the correct one. Structure expansion revealed the complete structure, which refined to a final R of 0.056. From this result the structure was unequivocally established as 2,6a,12-trimethyl-4,9-diisopropylperhydro-benzo[de]naphthacene (Fig. 4).

This structure indeed consists of two cadinane units; the four high-field protons observed in the ¹H-NMR spectrum are located at positions 1, 3, 8 and 13. Based on this identification and the similarity of the mass spectra of compounds W, R and T, we

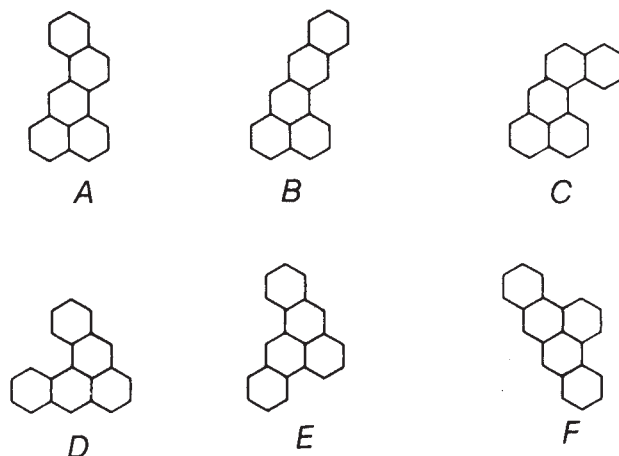


Fig. 3 Pentacyclic skeletons possible which contain 21 carbon atoms.

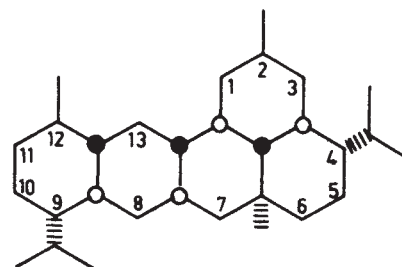


Fig. 4 Structure of compound T.

believe that compounds W and R may be stereoisomers of the compound identified. Further research is necessary to confirm this hypothesis.

We have encountered only one natural compound containing a dimeric cadinane skeleton; this is gossypol, a toxic yellow pigment from cotton seed oil¹³. However, its structure is completely different from that described above.

Several possibilities for the origin of this novel pentacyclic compound can be considered. The hydrocarbon could be derived from a functionalized precursor of biosynthetic origin. The biosynthesis could involve a cyclization of an acyclic C₃₀ isoprenoid compound or a dimerization of cadinene, known to occur in several plants. We emphasize, however, that biosynthesis by direct cyclization of squalene seems impossible in this case. A dimerization as the result of a purely chemical process in the geosphere cannot be ruled out completely at this stage.

We thank the Koninklijke/Shell Exploratie en Productie Laboratorium (Shell Research BV) for providing samples and for some financial support. We acknowledge the technical assistance of Mrs J. K. C. Hessels and Mrs E. H. Nieberg-van Velzen in isolating the compound.

Received 22 July; accepted 28 October 1985.

1. Grantham, P. J., Posthuma, J. & Baak, A. in *Advances in Organic Geochemistry 1981* (eds Bjorøy, M. *et al.*) 675-683 (Wiley, New York, 1983).
2. Egert, E. & Sheldrick, G. M. *Acta crystallogr.* **A41**, 262-268 (1985).
3. Hills, I. R. & Whitehead, E. V. in *Advances in Organic Geochemistry 1966* (eds Hobson, G. D. & Speers, G. C.) 89-110 (Pergamon, Oxford, 1970).
4. Aue, W. P., Bartholdi, E. & Ernst, R. R. *J. chem. Phys.* **64**, 2229 (1976).
5. Nagayama, K., Kumar, A., Wüthrich, K. & Ernst, R. R. *J. magn. Res.* **40**, 321 (1980).
6. Daneels, D. & Anteunis, M. *Org. magn. Reson.* **6**, 617-621 (1974).
7. Patt, S. L. *J. magn. Res.* **46**, 535 (1982).
8. Doddrell, D. M., Pegg, D. T. & Bendall, M. R. *J. magn. Res.* **48**, 323 (1982).
9. Bax, A. & Morris, G. *J. magn. Res.* **42**, 501 (1981).
10. Moshonas, M. G. & Lund, E. D. *Flavour Ind.* **1**, 375-378 (1970).
11. Van de Meent, D., Brown, S. C., Philp, R. P. & Simoneit, B. R. T. *Geochim. cosmochim. Acta* **44**, 999-1013 (1980).
12. Allinger, N. I. *J. Am. chem. Soc.* **99**, 8127-8134 (1977).
13. Adams, R., Morris, R. C., Geissman, T. A., Bufferbough, D. J. & Kirkpatrick, E. C. *J. Am. chem. Soc.* **60**, 2193-2204 (1938).

Early Eocene *Cantius torresi*—oldest primate of modern aspect from North America

Philip D. Gingerich

Museum of Paleontology, University of Michigan, Ann Arbor, Michigan 48109, USA

The oldest primates of modern aspect (primates of prosimian tarsiiform-lemuriform or higher grade) are known from Lower Eocene strata of Sparnacian age in Europe^{1,2}, Bumbanian age in Asia^{3,4}, and Wasatchian age in North America⁵. These first appearances coincide on all three northern continents with initial records of Artiodactyla, Perissodactyla and other modern mammalian groups appearing approximately 53 Myr BP^{6,7}. Here I report that a newly discovered basal Wasatchian faunal assemblage older than any known previously includes the well-preserved dental remains of a new species of *Cantius*, the earliest representative of Eocene Adapidae and the oldest evidence of primates of modern aspect known from North America. This new *Cantius* is possibly ancestral to all later adapids, including the European *Donrussellia* and North American *Pelycodus*. *Cantius torresi* has premolars reminiscent of those in Omomyidae, suggesting that earlier lemuriform and tarsiiform primates were closely related and little differentiated at the beginning of the Eocene.

Two major groups of primates of modern aspect, tarsiiform Omomyidae and lemuriform Adapidae, are represented in the early Eocene of Europe⁸⁻¹², Asia^{3,4,13} and North America¹⁴⁻²⁰. Both groups appeared suddenly on northern continents during an interval of climatic warming²¹. No plausible ancestors of the two groups are present in known European, Asian or North American faunas of late Palaeocene age, suggesting their immigration from equatorial areas. The oldest North American omomyid, *Teilhardina americana*, is similar to the European *Teilhardina belgica*^{15,19}. The oldest North American adapid known previously, *Cantius ralstoni*, is likewise similar to the European *Cantius eppsi*^{17,18}. The discovery of a new North American *Cantius* demonstrably older than *C. ralstoni* is important in clarifying the dental morphology of primitive Adapidae.

Cantius torresi n. sp.

Holotype. University of Michigan [UM] 83470, left dentary with P₃-M₁ collected by Victor Torres from locality SC-67 in the Willwood Formation at the south end of Polecat Bench, Park County, Wyoming.

Age and distribution. Earliest Wasatchian Land-Mammal Age (early Eocene) of the Clark's Fork Basin, Wyoming, western North America. This species is known from two adjacent localities (SC-67 and 69) in a restricted 10-m stratigraphic interval within the Clarkforkian-Wasatchian boundary sandstone of Kraus²².

Referred specimens. Holotype; UM 83467, left dentary with M₂; and UM 83475, right maxilla with M¹⁻³ from the type locality. UM 66143, right dentary with M₂₋₃ from SC-69.

Diagnosis. Differs from *C. ralstoni*, *C. eppsi* and all other species of *Cantius* in being smaller, having relatively square upper molars, and having shorter, relatively broader lower premolars and molars. P₄ is distinctive in having a broad trigonid and talonid. The paracristid connecting the paraconid and protoconid on P₄ forms a right-angle with the metacristid connecting the metaconid and protoconid.

Description. The holotype dentary (Fig. 1a) preserves matrix-filled alveoli for the lower canine, a single-rooted P₁ and a double-rooted P₂. The crown of P₃ is simple with a single central protoconid cusp, no accessory paraconid or metaconid, and little more than a trace of a talonid. P₃ is 2.4 mm long and

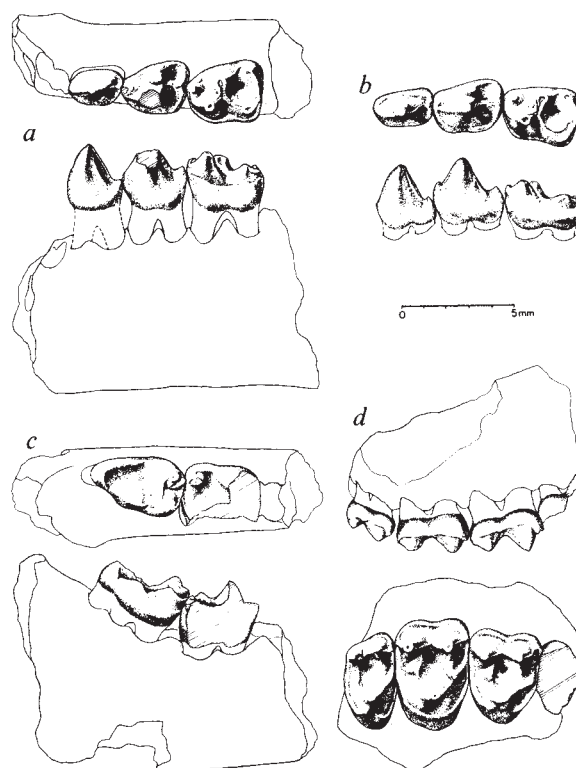


Fig. 1 Dentition of *Cantius torresi*, from the earliest Wasatchian of western North America, and *Cantius eppsi*, from the Sparnacian of UK. **a**, Left dentary of *C. torresi* [UM 83470, holotype] with P₃-M₁ in occlusal (top) and lateral (bottom) view. Partial alveoli for the single-rooted P₁ and double-rooted P₂ are shown in outline. The remains of the alveolus for the larger lower canine are obscured in these views. **b**, Right P₃-M₁ (reversed for comparison) of *C. eppsi* [Br. Mus. (Nat. Hist.) 15773, holotype] in occlusal and lateral view. **c**, Right dentary of *C. torresi* [UM 66143] with M₂₋₃ in occlusal and lateral view. **d**, Right maxilla of *C. torresi* with M¹⁻³ in lateral and occlusal view. Note the short *Nannopithecus* fold of enamel running up the posterior surface of the protocone in occlusal view. This enamel fold does not join the postcingulum, and the earliest *Cantius* lacks a fully developed postprotocingulum.

1.8 mm in maximum breadth. The crown of P₄ is more complex, with a well-developed paraconid and metaconid in addition to the central protoconid. The metaconid is medial to the protoconid, making the trigonid unusually broad. The talonid on P₄ is also broad, with a prominent central cristid obliqua and shallowly basined posteromedial corner. P₄ is 2.8 mm long and 2.4 mm in maximum breadth. The crown of M₁ is tritubercular with three distinct trigonid cusps and a broadly basined talonid. M₁ is 3.3 mm long and 2.9 mm in maximum breadth. The crown of M₂ is partially preserved in two specimens (UM 66143 and 83467), both are small, measuring 3.4 × 3.2 mm and 3.2 × 3.1 mm in length and width, respectively. The crown of M₃ preserved in UM 66143 has a short trigonid, with paraconid and metaconid cusps closely appressed (Fig. 1c). M₃ is 4.3 mm long and 2.7 mm in maximum breadth. The talonid of M₃ resembles that in other early *Cantius* species in having a very narrow hypoconulid lobe (1.5 mm broad).

Upper molars are preserved intact in one specimen (UM 83475; Fig. 1d). The crowns of both M¹ and M² are relatively square by comparison with later species of *Cantius*; both have prominent protocones, distinct paracones and metacones, and small but distinct paraconules and metaconules. There is a crest behind the protocrista representing a '*Nannopithecus* fold' or partially developed postprotocingulum as in other primitive adapids and omomyids. In later Adapidae this crest is either reduced and lost (*Donrussellia*, *Protoadapis*) or it is expanded

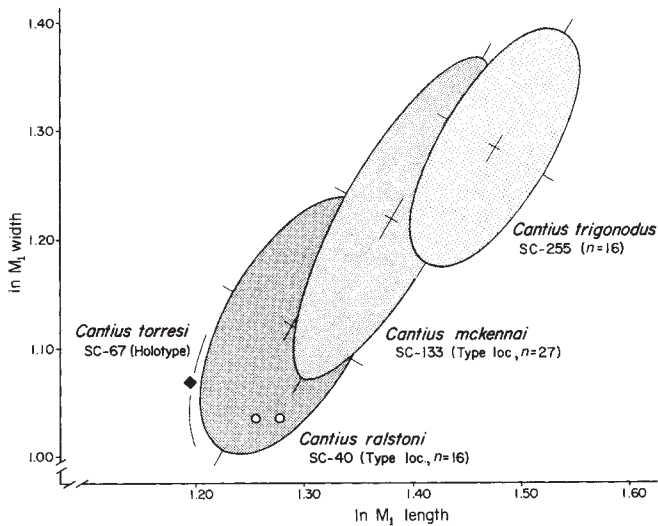


Fig. 2 Comparison of the sizes of the first lower molars in four species of *Cantius* from successive stratigraphically superposed intervals of the early and middle Wasatchian, Clark's Fork Basin, Wyoming. Equiprobability ellipses enclose 90% of the observed variation in early Wasatchian *C. ralstoni* (type sample), *C. mckennai* (type sample), and the later *C. trigonodus*. The holotype of *C. torresi* is just outside the 95% equiprobability contour for *C. ralstoni*. Open circles within the distribution of *C. ralstoni* represent the type sample of *C. eppsi* from Abbey Wood, Kent. Note that *C. torresi* differs from *C. ralstoni* and *C. eppsi* in shape as well as size, having a shorter and relatively broader M_1 .

into a full postprotocingulum (later *Cantius*, *Pelycodus*). There is a complete lingual cingulum on M^2 . M^1 is 3.3 mm long and 4.5 mm in maximum breadth. M^2 is 3.3 mm long and 5.3 mm in maximum breadth. The crown of M^3 is shorter and consequently relatively broader than those of M^1 and M^2 ; it is 2.5 mm long and 4.3 mm in maximum breadth.

Judging from the available evidence, *C. torresi* is the smallest and most generalized species of *Cantius* known anywhere. It is similar in size to *C. eppsi* from the Sparnacian Blackheath Beds of the United Kingdom (Abbey Wood, Kent), with a body weight estimated²³ from tooth size to be 1,100–1,200 g. *C. torresi* differs from *C. eppsi* in having shorter and broader premolars (Fig. 1a, b) and molars (Fig. 2). The type sample of *C. torresi* comes from the Clark's Fork Basin of northwestern Wyoming where more than 760 specimens of *Cantius*, including the type sample of *C. ralstoni* and the type sample of *Cantius mckennai*, are linked in one evolving lineage. As the oldest member of this lineage, *C. torresi* effectively defines the morphology of primitive North American *Cantius*.

Comparing the European *C. eppsi* from the type locality, Abbey Wood, with the North American sequence of *Cantius* species in the Clark's Fork Basin, the size and proportions of upper and lower cheek teeth agree most closely with *C. ralstoni* (Fig. 2), which suggests that *C. eppsi* and *C. ralstoni* represent equivalent grades of evolution and, further, that *C. eppsi* and *C. ralstoni* may be approximately equivalent in age. Radiometric dating and correlation based on palaeomagnetic stratigraphy are insufficiently precise to resolve time on this scale, but comparison of the fauna from SC-67, the type locality of *C. torresi*, with that from Abbey Wood substantiates an older age for *C. torresi*. Perissodactyls are particularly useful here, *Hyracotherium* from SC-67 being smaller and notably less lophodont than any European specimens (including problematical remains from the French locality of Meudon).

Donrussellia provincialis, a European contemporary of *C. eppsi*^{1,12}, differs from *C. torresi* in being smaller and having the anteroposteriorly elongated P_4 characteristic of *Protoadapis*

and later Adapidae, but it could be derived from a generalized *C. torresi* or a similar structural ancestor. *C. torresi* has relatively square upper molars like those of the later North American *Pelycodus jarrovi*^{18,24} and, if intermediates were known connecting them through time, *C. torresi* could easily be regarded as an early species of *Pelycodus*.

Of greater interest and importance is the fact that the form of P_4 in *C. torresi* is as similar to the fourth premolars of early tarsiiform Omomyidae as it is to those of later lemuriform Adapidae, suggesting that Omomyidae and Adapidae (and the prosimian infra-orders that they represent) were probably still little differentiated at the beginning of the Eocene. This premolar conformation linking Omomyidae and Adapidae serves, at the same time, to distinguish them from all archaic Palaeocene pleiadapiform primates of more primitive 'praesimian' grade²⁵. Similarly, the presence of a limited *Nannopithec* fold, rather than a full postprotocingulum connecting the protocone to the posterior cingulum, is another feature distinguishing primitive Adapidae and Omomyidae from earlier pleiadapiform primates with a fully developed postprotocingulum. This weakens earlier dental evidence suggesting close linkage of *Plesiadapis* and its allies in the order Primates²⁶. While I would still classify *Plesiadapis* as a primate²⁷, new evidence and interpretations presented here and elsewhere^{11,16,25,27,28} indicate that Plesiadapiformes represent an evolutionary radiation distinct from primates of modern aspect.

Previous comparisons of European Sparnacian mammalian faunas with known North American early Wasatchian faunas indicated that *C. eppsi* and contemporary European mammals of other orders were more primitive and therefore probably older than *C. ralstoni* and its North American contemporaries^{12,29}, suggesting that primates of modern aspect, Artiodactyla, hyaenodontid *Credonta*, Perissodactyla and other Wasatchian groups reached North America by dispersal from Europe. The discovery of a new and distinctive mammalian assemblage at the base of the North American Wasatchian, including a primate *C. torresi* that is more primitive than any known European adapid, complicates this interpretation. Based on present evidence, it seems that Sparnacian primates of modern aspect and at least some representatives of other modern mammalian orders reached Europe by dispersal through North America.

The geographic area of origin of primates of modern aspect is as yet unknown, but the presence of Lemuroidea on Madagascar today, Lorisioidea in Africa since the early Miocene³⁰, a possible Oligocene adapid (*Oligopithecus*) in Africa³¹, and the recent discovery of Oligocene Tarsiioidea in Africa³² make it likely that primates of prosimian grade originated on the African continent. Africa and Europe were separated by the Tethys Sea and Obik Sea throughout the Palaeocene and most of the Eocene³³. However, Africa may have been connected to central Asia by way of the Arabian peninsula and Indo-Pakistan subcontinent³⁴. Climatic warming in the late Palaeocene and earliest Eocene²¹ undoubtedly facilitated poleward dispersal of endemic equatorial African mammals. Thus, one may reasonably speculate that the earliest primates of modern aspect evolved in equatorial Africa during the late Palaeocene and dispersed rapidly through Arabia, South Asia, Central Asia and North America to reach Europe as part of a broader pattern of faunal interchange on northern continents leading to the holarctic cosmopolitanism evident in the early Eocene.

I thank T. M. Bown and R. D. Martin for reviewing the manuscript, and J. J. Hooker for casts and discussion. Specimens were prepared by W. Ryan, figures by K. Klitz. This research was supported by NSF grants BNS 80-16742, DEB 82-06242 and EAR 84-08647.

Received 10 May; accepted 1 November 1985.

- Godinot, M. *C. r. hebdom. Séanc. Acad. Sci., Paris D286*, 1869–1872 (1978).
- Hooker, J. J. & Insole, A. N. *Tertiary Res., Rotterdam* 3, 31–45 (1980).
- Dashzeveg, D. & McKenna, M. C. *Acta palaeont. pol.* 22, 119–137 (1977).

4. Dashzeveg, D. *Bull. Soc. géol. Fr. Sér. 7* **24**, 275-281 (1982).
5. Rose, K. D. *Science* **208**, 744-746 (1980).
6. Berggren, W. A., McKenna, M. C., Hardenbol, J. & Odradovich, J. D. *J. Geol.* **86**, 67-81 (1978).
7. Harland, W. B. et al. *A Geologic Time Scale* (Cambridge University Press, 1982).
8. Simons, E. L. *Bull. Br. Mus. nat. Hist., Geol.* **7**, 1-36 (1962).
9. Russell, D. E., Louis, P. & Savage, D. E. *Univ. Calif. Publ. Geol. Sci.* **73**, 1-46 (1967).
10. Szalay, F. S. *Folia Primatol.* **22**, 116-133 (1974).
11. Gingerich, P. D. *Folia Primatol.* **28**, 60-80; 144-153 (1977); *J. hum. Evol.* **10**, 345-374 (1981).
12. Godinot, M. *Palaeovertebrata* **10**, 43-126 (1981); *Geobios, Mém. Spéc.* **6**, 403-412 (1982).
13. Rose, K. D. & Krause, D. W. *J. Mammal.* **65**, 721-726 (1984).
14. Russell, D. E. & Gingerich, P. D. *C. r. hebdom. Séanc. Acad. Sci., Paris D291*, 621-624 (1980).
15. Bown, T. M. *Folia Primatol.* **25**, 62-72 (1976).
16. Szalay, F. S. *Bull. Am. Mus. nat. Hist.* **156**, 157-450 (1976).
17. Gingerich, P. D. & Simons, E. L. *Contr. Mus. Paleont. Univ. Mich.* **22**, 245-279 (1977).
18. Gingerich, P. D. & Haskin, R. A. *Contr. Mus. Paleont. Univ. Mich.* **25**, 327-337 (1981).
19. Rose, K. D. & Bown, T. M. *Nature* **309**, 250-252 (1984).
20. Bown, T. M. & Rose, K. D. *Folia Primatol.* **43**, 97-112 (1984).
21. Buchardt, B. *Nature* **275**, 121-123 (1978).
22. Kraus, M. J. in *Early Cenozoic Paleontology and Stratigraphy of the Bighorn Basin, Wyoming* (ed. Gingerich, P. D.) 87-94 (University of Michigan Museum of Paleontology, Ann Arbor, 1980).
23. Gingerich, P. D., Smith, B. H. & Rosenberg, K. *Am. J. phys. Anthropol.* **58**, 81-100 (1982).
24. Rose, K. D. & Bown, T. M. *J. Paleont.* **58**, 1532-1535 (1984).
25. Gingerich, P. D. *Yb. phys. Anthropol.* **27**, 57-72 (1984).
26. Gingerich, P. D. in *Prosimian Biology* (eds Martin, R. D., Doyle, G. A. & Walker, A. C.) 531-541 (Duckworth, London, 1980).
27. Gingerich, P. D. in *Major Topics in Primate and Human Evolution* (eds Wood, B. A., Martin, L. B. & Andrews, P.) (Cambridge University Press, in the press).
28. MacPhee, R. D. E., Cartmill, M. & Gingerich, P. D. *Nature* **301**, 509-511 (1983).
29. Gingerich, P. D. *A. Rev. Earth planet. Sci.* **8**, 407-424 (1980).
30. Walker, A. C. in *Prosimian Biology* (eds Martin, R. D., Doyle, G. A. & Walker, A. C.) 435-447 (Duckworth, London, 1980).
31. Gingerich, P. D. *Geobios, Mém. Spéc.* **1**, 165-182 (1977).
32. Simons, E. L. & Bown, T. M. *Nature* **313**, 475-477 (1984).
33. McKenna, M. C. *Ann. M. bot. Gdn* **62**, 335-353 (1975).
34. Smith, A. G. & Briden, J. C. *Mesozoic and Cenozoic Paleogeographic Maps* (Cambridge University Press, 1977).

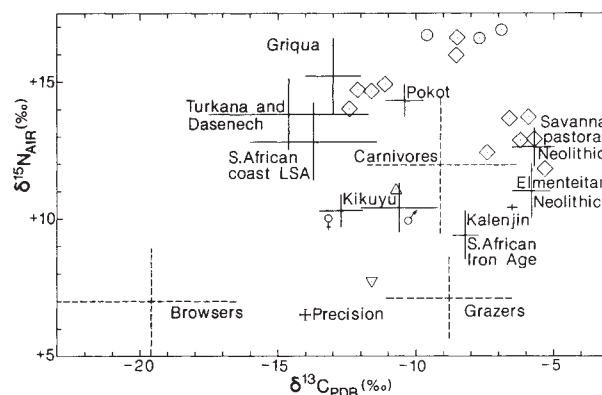


Fig. 1 $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ values of bone collagen from African human and non-human mammal populations and individuals. The means (± 1 s.d.) are indicated by symbols or solid lines for human groups and by broken lines for modern non-human East African mammals. The symbols represent individual determinations for the following human groups and individuals: Δ , proto-historic Kikuyu from site GsJj31, Naivasha, Kenya; ∇ , proto-historic Luyia from site GnJc51, Bungoma, Kenya; $\diamond$ and $\diamond$, prehistoric individuals from Sechikuencho Cairns and from Gishimangeda Cave respectively, Lake Eyasi, northern Tanzania.

Reconstruction of African human diet using bone collagen carbon and nitrogen isotope ratios

Stanley H. Ambrose* & Michael J. DeNiro†

Department of Earth and Space Sciences, and †Archaeology Program, University of California, Los Angeles, California 90024, USA

Behavioural modifications associated with the exploitation of new food resources have been linked to major steps in hominid evolution and in subsequent human cultural development^{1,2}. Testing of specific hypotheses concerning the influence of dietary change on these processes would be facilitated by quantitative estimates of early hominid and human diets. Although most methods of obtaining such evidence provide only qualitative information², the stable carbon and nitrogen isotope ratios of animal tissues, and in particular, bone collagen, can be used to quantify the consumption of foods having different isotopic compositions³⁻⁵. As reported here, analysis of the collagen of historic and prehistoric African human populations from Kenya, Tanzania and South Africa that have reasonably well-known diets shows that stable carbon and nitrogen isotope ratios in bone collagen can distinguish marine foragers from populations consuming terrestrial resources, pastoralists from farmers, farmers consuming grains from those consuming non-grain crops, and camel pastoralists from capri-bovine pastoralists.

Several major classes of food resources can be distinguished on the basis of their isotopic composition. Among terrestrial plants, carbon isotope ratios differentiate between C_3 plants and C_4 and CAM (crassulacean acid metabolism) plants⁶. C_4 plants include maize, sorghum, African millets and tropical pasture grasses. Most succulents such as cacti are CAM plants. Other staple foods, including wheat, rice, nuts and most vegetables and fruits, are C_3 plants⁶. The average $\delta^{13}\text{C}$ values (see Table

1 for an explanation of δ values) for C_3 , C_4 and CAM, and marine plants, are -26.5 , -12.0 , and -19.0 ‰, respectively⁶. Nitrogen isotope ratios distinguish marine from terrestrial plants and nitrogen-fixers from other plants. The $\delta^{15}\text{N}$ values of marine plants are ~ 4 ‰ higher than those of terrestrial plants that use soil nitrogen⁵. Plants that use fixed atmospheric nitrogen (for example, legumes) have lower $\delta^{15}\text{N}$ values than other plants⁷. The isotopic compositions of animal tissues reflect the isotope ratios of the plants at the bases of food chains^{3,4}. However, there is a small stepwise enrichment of ^{13}C and a larger one of ^{15}N as carbon and nitrogen move to higher trophic levels³⁻⁵. In fact, enrichment in ^{15}N is large enough to serve as an indicator of trophic level⁵.

Given these basic differences in the isotopic composition of food resources, it should be possible to use $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values to distinguish between human populations relying mainly on animal products as opposed to plant foods, the meat of grazing animals as opposed to that of browsing animals, or C_3 as opposed to C_4 plants. These expectations are fulfilled for modern non-human mammals collected in Kenya and Tanzania⁸ (see Fig. 1). These results, based on the analysis of 15 carnivores, 111 grazers and 82 browsers (10, 13 and 18 species, respectively) serve as a standard against which isotopic data for humans presented in this study can be evaluated.

Bones of historic, proto-historic (Table 1) and prehistoric (Table 2) humans were obtained from museum collections in Kenya, Tanzania and South Africa. Bone collagen was purified and its $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values determined by mass spectrometry, as described elsewhere³⁻⁵.

Table 1 lists, in rank order of importance, the major components of the diets of Kikuyu farmers of central Kenya⁹, Luyia farmers of western Kenya¹⁰, Turkana and Dasenech pastoralists of northern Kenya^{11,12}, highland and lowland Pokot of north-west Kenya^{13,14}, Nandi and Kipsigis (Kalenjin) farmers and pastoralists of highland western Kenya¹³, and Grikua pastoralists and farmers of the Orange Free State, South Africa¹⁵.

The diets of the prehistoric populations are less well understood because of food preservation biases inherent in the archaeological record. The diet of Elmenteitan Neolithic and Iron Age peoples of highland western Kenya is considered to have been similar to that of the historic Kalenjin¹⁶. The diet of Savanna Pastoral Neolithic (SPN) peoples appears to have been dominated by cattle and caprine pastoralism but their degree of dependence on agriculture is under debate^{16,17}. There is inadequate information regarding the diets, ages and cultural

* Present address: Department of Anthropology, University of Illinois, Urbana, Illinois 61801, USA.

Table 1 $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ values (‰) of bone collagen from historic and proto-historic human populations from Kenya, Tanzania and South Africa

Population	$\delta^{15}\text{N}$				$\delta^{13}\text{C}$				No. of individuals	Diet	Ref.
	$\bar{x}$	s.d.	Min.	Max.	$\bar{x}$	s.d.	Min.	Max.			
Kenya											
Kikuyu (δ)	+10.4	0.9	+9.3	+12.3	-10.6	1.4	-13.1	-7.3	12	C_3/C_4 agriculture, with small amounts of cattle and caprines	11
Kikuyu (φ)	+10.3	0.6	+9.6	+11.2	-12.7	0.8	-13.9	-11.6	5		
Kikuyu, 19th century (GsJj31, Naivasha)	+11.1				-10.7				1		
Turkana/Dasenech	+13.8	1.3	+12.1	+16.0	-14.6	2.9	-18.0	-10.4	7	Camel, cattle, caprine pastoralism, some C_4 grains and wild C_3 plants	13, 14
Pokot	+14.2		+13.4	+14.9	-10.4		-11.2	-9.7	2	Cattle, caprine and camel pastoralism and C_4 grains	15, 16
Kalenjin	+10.4		+10.3	+10.5	-6.5		-6.7	-6.3	2	C_4 grains, and cattle and caprine pastoralism	15
Luyia, 19th century (GnJc51, Bungoma)	+7.7				-11.6				1	C_3/C_4 agriculture with small amounts of cattle and caprines	12
South Africa											
Griqua, 19th century	+15.2	1.4	+13.5	+17.7	-13.0	1.0	-14.7	-11.4	12	Cattle and caprine pastoralism and C_3 agriculture	17

Major classes of food resources are listed in rank order of importance in the diet of the populations. Isotope ratios are expressed as δ values, where $\delta^*X = [((X/X)_{\text{sample}}/(X/X)_{\text{standard}}) - 1] \times 1,000\text{‰}$. For carbon, X/X is $^{13}\text{C}/^{12}\text{C}$ and the standard is the PDB carbonate. For nitrogen, X/X is $^{15}\text{N}/^{14}\text{N}$ and the standard is atmospheric (Ambient Inhalable Reservoir, AIR) nitrogen.

affiliations of the individuals from the Eyasi Basin, Northern Tanzania¹⁸. The cultural and economic history of this region is among the most complex in Africa¹⁸, so it is possible that not all the prehistoric individuals had the same diets. The diet of the Later Iron Age peoples of the Northern Transvaal, South Africa, is considered to have been similar to that of the historic Pedi, who depend mainly on C_4 grains, supplemented by wild and domestic C_3 plants; cattle, caprines and wild animals contribute little to their diet¹⁹. The Later Stone Age hunter-gatherers of the South African Cape coast between Port Elizabeth and Hermanus subsisted on marine vertebrates and invertebrates, terrestrial C_3 plants and the flesh of herbivores²⁰.

The results of the isotopic analyses of the historic populations can be compared with published descriptions of their diets (Table 1, Fig. 1). Those groups described as being most dependent on plant foods (Luyia, Kikuyu and Kalenjin) have $\delta^{15}\text{N}$ values between +8 and +12‰, while those most dependent on the milk, meat and blood of domestic animals (Turkana, Dasenech, Pokot and Griqua) have $\delta^{15}\text{N}$ values between +12 and +18‰. The Kalenjin, who consume mostly C_4 plants and animals that grazed on C_4 grasses, have a mean $\delta^{13}\text{C}$ value of -6.5‰. Those populations depending on a mix of C_3 and C_4 plants and/or a mix of products from C_4 -grazing and C_3 -browsing animals (Luyia, Kikuyu, Pokot, Turkana, Dasenech and Griqua) have mean $\delta^{13}\text{C}$ values ranging from -10 to -15‰.

The pastoral/agricultural and C_3/C_4 distinctions are also apparent among the prehistoric populations (Table 2, Fig. 1). The mean $\delta^{15}\text{N}$ values for the South African Later Iron Age, Elmenteitan and SPN populations are +9.4, +11.0 and +12.7‰, respectively. This ranking by $\delta^{15}\text{N}$ values is consistent with archaeological evidence for their degree of dependence on animal as opposed to plant products. The Elmenteitan and SPN groups have virtually identical mean $\delta^{13}\text{C}$ values, indicating nearly exclusive use of C_4 -based resources; however, $\delta^{15}\text{N}$ values indicate that they fed at different trophic levels. The mean $\delta^{13}\text{C}$ value of -8.2‰ for the South African Later Iron Age population indicates their relatively greater dependence on C_3 plant foods relative to East African Neolithic populations.

The prehistoric southern Cape coastal forager sample set (Table 2) has a mean $\delta^{15}\text{N}$ value of +12.8‰, which is within the range of previously analysed populations from other coastal

regions²¹. Their $\delta^{13}\text{C}$ values, with one exception, reflect a diet of no more than 50% terrestrial resources²¹. As in the western Cape²², our isotopic evidence seems inconsistent with a model of seasonal transhumance between coastal and montane environments²⁰. However, this region is transitional between winter and summer rainfall regions so the grass flora at different locations within the region contains from 0 to 75% C_4 grasses²³. Therefore, skeletons from montane environments need to be analysed in order to confirm this conclusion.

The $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values of the 15 prehistoric individuals from Tanzania¹⁸ form three discrete clusters (Fig. 1). Six individuals are within the range of the Kenyan SPN. The second cluster has mean $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ values of +14.6 and -11.8‰, respectively, indicating a diet dominated by animal resources that included some browsing animals, supplemented by wild plant foods and/or C_3 crops. The third cluster has the highest average $\delta^{15}\text{N}$ value in the sample set and must reflect a diet dominated by the protein of grazing animals. The $\delta^{13}\text{C}$ values for this cluster indicate that small amounts of C_3 plants and/or browsing animals were eaten.

The agreement between the isotopic composition of bone collagen from the populations analysed and their diets as inferred from ethnographic and archaeological evidence is remarkably close in almost every case. Those groups depending largely on the milk, meat and blood of domestic animals have the highest $\delta^{15}\text{N}$ values, while those that subsist mainly on plants have the lowest ones. Groups subsisting mainly on C_4 plants or the protein of grazing herbivores have higher $\delta^{13}\text{C}$ values than those depending on a mix of C_3 and C_4 plants and grazing and browsing animals.

There is also close agreement between the isotopic compositions of prehistoric and historic members of the same ethnic groups. The nineteenth-century Kikuyu individual has an isotopic composition that is within the range determined for twentieth-century Kikuyu males (Table 1, Fig. 1). It has been proposed that the Kalenjin are the descendants of Elmenteitan Neolithic peoples^{16,24} and the similarity in the isotopic compositions suggests similar diets.

The isotopic composition of the Kenyan Neolithic groups helps us to resolve the current controversy over their dietary adaptations^{16,17}. Evidence from linguistic studies suggests that

Table 2 $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ values (‰) of bone collagen from humans recovered from archaeological sites in Kenya, Tanzania and South Africa

Group and location	Site	$\delta^{15}\text{N}$	$\delta^{13}\text{C}$	Ref.
Elmenteitan	Deloraine Farm (GqJh6)	+10.6	-4.8	24
Neolithic and Iron Age, Kenya	Rigo Cave (GrJh3/44)	+11.8	-5.4	32
	Rigo Cave (GrJh3/32)	+10.3	-6.0	32
	Rigo Cave (GrJh3/33)	+10.9	-5.0	32
	Rigo Cave (GrJh3/6)	+11.7	-7.3	32
	Egerton Cave (GrJh10/2)	+11.4	-5.1	33
	Egerton Cave (GrJh10/3)	+12.2	-5.8	33
	Egerton Cave (GrJh10/4)	+10.9	-5.8	33
	Egerton Cave (GrJh10)	+8.6	-6.1	33
	Keringet Caves	+11.3	-6.4	34
Savanna Pastoral Neolithic (SPN), Kenya	Marula rockshelter (GSJ24)	+12.9	-6.5	35
	Cole's Farm (GrJ5a)	+11.4	-5.0	16
	Cole's Farm (GrJ5a/132)	+12.4	-4.8	16
	Lukunya Hill (GvJm202/541)	+12.0	-6.2	35
	Lukunya Hill (GvJm202/785)	+11.6	-5.1	35
	Lukunya Hill (GvJm202/513)	+13.1	-4.5	35
	Mt Suswa (GuJj14)	+13.1	-6.5	35
	Naivasha Burial Site (KNM-NV43)	+12.6	-6.9	36
	Kisima Farm (KFR-A5, no. 11)	+13.9	-5.9	37
	Makalia Burial Site (II)	+13.0	-5.9	37
Prehistoric burials, Lake Eyasi, Tanzania	Gishimangeda Cave	+12.9	-6.2	18
	Gishimangeda Cave	+16.6	-8.5	18
	Gishimangeda Cave	+11.8	-5.3	18
	Gishimangeda Cave	+12.4	-7.4	18
	Gishimangeda Cave (A)	+12.9	-5.7	18
	Gishimangeda Cave (B)	+15.9	-8.5	18
	Gishimangeda Cave (C)	+14.7	-11.6	18
	Gishimangeda Cave (D)	+14.0	-12.4	18
	Gishimangeda Cave (E)	+13.7	-5.9	18
	Gishimangeda Cave (F)	+14.7	-12.2	18
	Gishimangeda Cave (G)	+14.9	-11.2	18
	Gishimangeda Cave	+13.6	-6.6	18
	Sechikuencho Cairns	+16.9	-6.9	18
	Sechikuencho Cairns	+16.6	-7.7	18
	Sechikuencho Cairns	+16.7	-9.6	18
Later Iron Age, northern Transvaal, South Africa	Makapan (2429AA4)	+9.7	-7.9	38
	Maxonwa Maone (2329CD12)	+9.1	-8.3	38
	Bambo (2329CD9C)	+11.1	-8.6	38
	Tauhatswana	+8.0	-7.8	38
	(2329BB2A, no. 1)			
	Tauhatswana	+8.3	-7.8	38
	(2329BB2A, no. 2)			
	Bambo (2329CD9A, no. 3)	+9.4	-9.3	38
	Bambo (2329CD9A, no. 1)	+9.7	-8.0	38
	Nylsvley (2248DA)	+10.2	-7.8	38
Later Stone Age, southern Cape coast, South Africa	Knysna Strandlooper (A1070A)	+11.9	-14.5	39
	Knysna Strandlooper (A1070B)	+15.7	-11.2	39
	Schoenmakaskop (A1116)	+12.4	-13.2	39
	Loeritank (A1117)	+14.0	-12.0	39
	Sunday's River (A1118)	+14.0	-13.6	39
	Sea View (A1138)	+11.8	-11.0	39
	Zitzikama (A1155)	+13.2	-19.2	39
	Zitzikama Rock Shelters (A1112)	+13.3	-12.5	40
	Zitzikama Rock Shelters (A1112, ZA2)	+12.9	-15.5	40
	Byneskranskop 3 (no. 10)	+10.2	-14.6	41
	Byneskranskop 3 (no. 3)	+13.3	-11.4	41
	Byneskranskop 3 (no. 4)	+11.6	-15.4	41

both groups were familiar with C_4 grain crops²⁵, but archaeological evidence suggests that SPN populations were pastoralists rather than farmers¹⁶. The mean $\delta^{15}\text{N}$ values of the Elmenteitan and SPN groups are significantly different ($P < 0.01$), but both are intermediate between historic pastoral and agricultural groups. Lower $\delta^{15}\text{N}$ values for the Elmenteitan suggest an even mix of plant and animal resources whereas higher values for the SPN suggest a greater reliance on animal products. High $\delta^{13}\text{C}$ values indicate that the plant and animal components of the diets of both groups were almost entirely C_4 grains and grazing animals.

The $\delta^{15}\text{N}$ values of the pastoral populations are significantly higher than those of the East African carnivores (Fig. 1). This

unexpected result suggests that the former populations consumed resources having $\delta^{15}\text{N}$ values higher than those eaten by the carnivores. Milk and blood appear to be enriched in ^{15}N relative to the flesh of animals²⁶, which could account for this observation.

Our results suggest that it should be possible to use stable isotopes to test models of prehistoric hunter-gatherer diets. For example, most modern hunter-gatherers in the tropics live in marginal environments and are heavily dependent on plant foods²⁷. However, before the advent of food production they inhabited non-marginal environments and may have exploited fewer plant foods²⁸. If this dietary shift can be demonstrated isotopically, widely held concepts of tropical hunter-gatherer adaptations, and of the relative significance of hunting versus gathering in human evolution, would require revision.

We can also speculate on the expected isotopic composition of different Plio-Pleistocene hominids. Australopithecine diets are considered to have been predominantly vegetarian²⁹. As virtually all wild plant foods now eaten by African humans and primates³⁰ are C_3 and the consumption of C_4 grasses can be ruled out on the basis of dental microwear evidence², we would expect Australopithecines to have low $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values. Archaeological evidence suggests that *Homo habilis* and *Homo erectus* consumed the meat of grazing animals³¹, so we would expect them to have higher $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ values than the Australopithecines. However, before early hominids are analysed it will be necessary to demonstrate that organic fractions retaining their *in vivo* isotopic record exist in samples of Plio-Pleistocene age and that stable isotopes can distinguish among grazers, browsers and carnivores contemporary with the early hominids.

We thank the National Museums of Kenya and Tanzania, the South African Museum and the Departments of Archaeology and Anatomy, University of Witwatersrand, and numerous individuals for providing human bone samples. Support was provided by the L.S.B. Leakey Foundation, The Foundation for Research into the Origin of Man, and NSF grants BNS79-24756, BNS81-18026 and BNS84-18280.

Note added in proof: All prehistoric samples discussed here had atomic carbon-to-nitrogen ratios in the range 3.0–3.5 and thus retain their *in vivo* isotopic signatures, according to the criterion proposed by DeNiro⁴².

Received 26 July; accepted 24 October 1985.

- Isaac, G. L. In *Evolution from Molecules to Men* (ed. Bendall, D. S.) 509–543 (Cambridge University Press, 1983).
- Walker, A. *Phil. Trans. R. Soc. B292*, 57–64 (1981).
- DeNiro, M. J. & Epstein, S. *Geochim. cosmochim. Acta* **42**, 495–506 (1978).
- DeNiro, M. J. & Epstein, S. *Geochim. cosmochim. Acta* **45**, 341–351 (1981).
- Schoeninger, M. J. & DeNiro, M. J. *Geochim. cosmochim. Acta* **48**, 625–639 (1984).
- Smith, B. N. *BioScience* **22**, 226–231 (1972).
- Delwiche, C. C., Zinke, P., Johnson, C. M. & Virginia, R. A. *Bot. Gaz.* **140** (Suppl.), 564–569 (1979).
- Ambrose, S. H. & DeNiro, M. J. *Oecologia* (submitted).
- Leakey, L. S. B. *The Southern Kikuyu Before 1903* Vol. 1 (Academic, London, 1977).
- Wagner, G. *The Bantu of North Kavirondo* Vol. 2 (International African Institute, London, 1956).
- Gulliver, P. & Gulliver, P. H. *The Central Nilo-Hamites* (International African Institute, London, 1953).
- Carr, C. J. *Pastoralism in Crisis: The Dasenech and their Ethiopian Lands* (Univ. Chicago Dep. Geogr. Res. Pap. 180, 1977).
- Huntingford, G. W. B. *The Southern Nilo-Hamites* (International African Institute, London, 1953).
- Conant, F. P. *Am. Anthropol.* **67**, 429–434 (1965).
- Morrison, A. thesis, Univ. Witwatersrand, Johannesburg (1984).
- Ambrose, S. H. in *From Hunters to Farmers* (eds Clark, J. D. & Brandt, S. A.) 212–239 (University of California Press, 1984).
- Robertshaw, P. T. & Collett, D. P. *Wild Archaeol.* **15**, 67–78 (1983).
- Ikeda, J. & Hayama, S. *Afr. Stud. Monogr.* **2**, 1–26 (1982).
- Mönnig, H. O. *The Pedi* (Van Schaik, Pretoria, 1967).
- Deacon, H. J. S. *Afr. archaeol. Soc. Goodwin Ser.* **1**, 26–45 (1972).
- Schoeninger, M. J., DeNiro, M. J. & Tauber, H. *Science* **220**, 1381–1383 (1983).
- Sealy, J. C. & van der Merwe, N. J. *Nature* **315**, 138–140 (1985).
- Vogel, J. C., Fuls, A. & Ellis, R. P. *S. Afr. J. Sci.* **74**, 209–215 (1978).
- Ambrose, S. H. *Azania* **19**, 79–104 (1984).

25. Ehret, C. *Ethiopians and East Africans: The Problem of Contacts* (East African Publishing House, Nairobi, 1974).
26. Steele, K. W. & Daniel, R. M. *J. agric. Sci.* **90**, 7-9 (1978).
27. Lee, R. B. in *Man the Hunter* (eds Lee, R. B. & DeVore, I.) 30-48 (Aldine, Chicago, 1968).
28. Ambrose, S. H. *Sugia (Sprache und Geschichte in Afrika)* **7** (in the press).
29. Pilbeam, D. & Gould, S. J. *Science* **186**, 892-901 (1974).
30. Peters, C. R. & O'Brien, E. M. *Curr. Anthropol.* **22**, 127-140 (1981).
31. Isaac, G. L. *Adv. Wild Archaeol.* **3**, 1-87 (1984).
32. Wandibba, S. *Azania* **18**, 81-92 (1983).
33. Faugust, P. M. & Sutton, J. E. *G. Azania* **1**, 149-175 (1966).
34. Cohen, M. *Azania* **5**, 27-38 (1970).
35. Ambrose, S. H. thesis, Univ. California, Berkeley (1984).
36. Leakey, L. S. B. *The Stone Age Races of Kenya* (Oxford University Press, London, 1935).
37. Sirinäinen, A. *Azania* **12**, 161-186 (1977).
38. Moore, M. J. thesis, Univ. Witwatersrand, Johannesburg (1981).
39. Fitzsimmons, F. W. S. *S. Afr. J. Sci.* **25**, 448-450 (1928).
40. Laing, G. D. & Gear, H. S. *S. Afr. J. Sci.* **26**, 575-601 (1929).
41. DeVilliers, H. & Wilson, M. L. *Ann. S. Afr. Mus.* **88**, 205-248 (1982).
42. DeNiro, M. J. *Nature* **317**, 806-809 (1985).

Blue light-dependent proton extrusion by guard-cell protoplasts of *Vicia faba*

K. Shimazaki*†, M. Iino*‡ & E. Zeiger†

† Department of Biological Sciences, Stanford University, Stanford, California 94305, USA

‡ Carnegie Institution of Washington, Department of Plant Biology, 290 Panama Street, Stanford, California 94305, USA

In plant leaves, light regulation of stomatal movement, and hence of gas exchange¹, involves a non-photosynthetic photosensory system sensitive to blue light²⁻⁶, which transduces many other plant movement responses. The exact nature of the primary photo-reactions of this system is still not known⁷. Stomatal opening is, in general, based on swelling of the guard cells driven by ion uptake¹, and it has been suggested that energy for such uptake may be provided by a chemi-osmotic extrusion of protons across the plasmalemma^{1,8,9}. We report here that guard-cell protoplasts from *Vicia faba*, irradiated continuously with photosynthetically saturating, high-fluence-rate red light, extrude H⁺ ions in response to a short pulse (30 s) of blue light. The response is relatively long-lived, such that a high rate of H⁺ extrusion continues for several minutes after the pulse stimulation, indicating that continuous input of blue light energy is not required for a sustained response. The kinetic relation between pulse input and response output matched closely that found for the stomatal response in intact leaves⁶. In agreement with the chemi-osmotic hypothesis, activation of an electrogenic plasmalemma H⁺-ATPase is postulated as the cause of blue light-induced H⁺ extrusion.

Proton transport in guard cells can be studied most effectively by using protoplasts because they can be obtained free of other contaminating cell types^{3,10,11} and because no highly buffering cell wall is present¹². The blue light-sensitive system can be effectively isolated for study with a dual-beam protocol in which a blue light pulse is applied during a continuous, background irradiation with high-fluence-rate red light. This method has been used successfully to study the stomatal opening transduced by the blue light-sensitive photosystem, without additional photosynthesis-related stomatal responses.

In dark-adapted preparations, red light irradiation caused an alkalinization of the medium (Fig. 1A) which lasted for about 1 h. (This response, frequently observed with photosynthetically competent cell preparations, is usually a consequence of CO₂ fixation¹³.) A 30-s pulse of blue light, given after medium pH reached a steady state under red light, caused an acidification that continued for about 10 min (Fig. 1B). Although the magni-

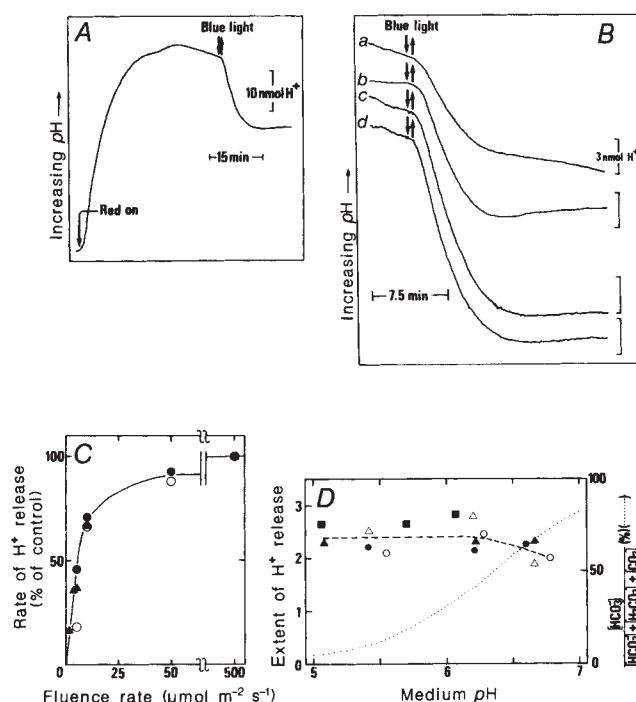


Fig. 1 Blue light-induced medium acidification by *Vicia* guard-cell protoplasts. **A**, Medium alkalinization in response to continuous red light (fluence rate, 800 $\mu\text{mol m}^{-2} \text{s}^{-1}$) and acidification in response to a subsequent blue light pulse (30 s, indicated by arrows; fluence rate, 500 $\mu\text{mol m}^{-2} \text{s}^{-1}$). Without a blue light pulse, the red light-induced alkalinization was followed by a nearly steady pH level. **B**, Typical acidification profiles in response to different fluences of blue light (pulse duration, 30 s), applied about 1 h after the onset of red light irradiation. *a*, 150; *b*, 300; *c*, 1,500; and *d*, 15,000 $\mu\text{mol m}^{-2}$ of blue light. Fluence rate was adjusted with neutral density filters. **C**, Rate of H⁺ release as a function of fluence rate of blue light (pulse duration, 30 s) in three different experiments. Rates were normalized relative to those measured in response to 500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ blue light. Identical symbols correspond to the same protoplast batch. **D**, Blue light-induced acidification as a function of medium pH, expressed as $\mu\text{mol H}^+$ released per mg chlorophyll per pulse. Identical symbols indicate measurements with the same protoplast batch. Dotted line is the percentage of HCO₃⁻ concentration in the total concentration of HCO₃⁻, CO₂ and H₂CO₃ in water, calculated from the Henderson-Hasselbach equation ($pK_1 = 6.35$, 25 °C).

Methods. Guard-cell protoplasts were isolated enzymatically from abaxial epidermis of 4- to 5-week-old leaves of *Vicia faba* L. (cv. Long Pod) as described previously¹¹. Suspension of isolated protoplasts (about 10⁶ cells) in 1.2 ml of 0.5 mM methane ethane sulphate-NaOH buffer (pH 6.2, unless otherwise stated) containing 0.35 M mannitol, 1 mM CaCl₂ and 10 mM KCl, was continuously stirred in an open glass vessel enclosed in a Plexiglas water jacket filled with circulating water at 25 °C. Medium pH was measured with a pH glass electrode (Beckmann 39522) connected to a pH meter (Radiometer 26) and a chart recorder. An intrinsic response of the pH electrode was eliminated by shielding it with a piece of Cinemoid 5A film. Amounts of acid equivalents were determined by the addition of 10 nmol H⁺ at the end of each experiment. (Blue light-induced acidification occurred within 0.05 pH units.) The chlorophyll concentration in each protoplast preparation was determined by the method of Arnon²². Rates of H⁺ release were calculated per unit weight of chlorophyll. Rates of H⁺ release per unit area of plasmalemma were estimated given an average cell chlorophyll content²¹ of 2 pg and an average protoplast diameter of 16 μm . Red light was obtained by passing light from a 500-W Sylvania tungsten lamp through a Corning 2-61 red filter and a layer of Cinemoid 5A; blue light, by passing light from a Kodak Carousel 4200 projector (300-W EXR lamp) through a Corning 5-60 filter and a 4-cm layer of a 0.2% copper sulphate solution. Fluence rates were measured with a Li-Cor quantum sensor.

* Present addresses: Division of Environmental Biology, National Institute for Environmental Studies, Yatabe-machi, Ibaraki 305, Japan (K.S.); Department of Biology, Tokyo Metropolitan University, Fuzakawa, Tokyo 158, Japan (M.I.).

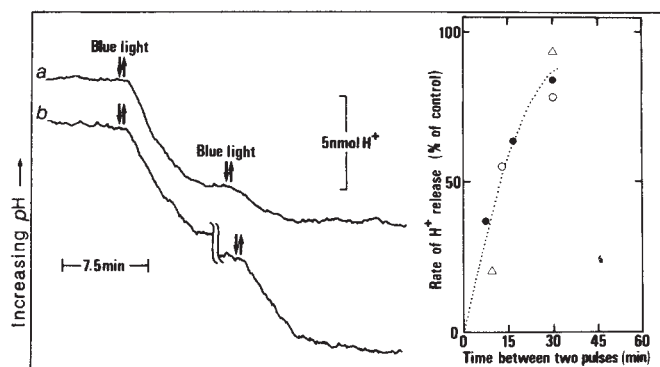


Fig. 2 H^+ release in response to two consecutive, saturating pulses (each pulse: 30 s with $50 \mu\text{mol m}^{-2} \text{s}^{-1}$). Traces *a* and *b*, two pulses given 7.5 and 30 min apart, respectively. Inset: rates of H^+ release in response to a second pulse, as a function of the time between two pulses. The rate of H^+ release in response to a second pulse is expressed as a percentage of that obtained with the first pulse. Background red light irradiation and other experimental details are as in Fig. 1. Identical symbols correspond to the same protoplast batch.

tude of the response was fluence-dependent, the relative time courses obtained with different fluences were nearly congruent (Fig. 1*B*), with maximal rates of acidification observed within 2 min after pulse stimulation. A pulse of red light (30 s with $500 \mu\text{mol m}^{-2} \text{s}^{-1}$) superimposed on background red irradiation, induced no detectable pH change, demonstrating that the acidification was a specific response to blue light.

The pulse-induced response described above was saturable (Fig. 1*C*), with half-saturations at about $180 \mu\text{mol m}^{-2}$ (30 s with $6 \mu\text{mol m}^{-2} \text{s}^{-1}$). The protoplasts once saturated by a single pulse, however, became responsive to another saturating pulse. This was indicated by an additional acidification caused by a second pulse given with sufficient delay (Fig. 2). These properties of the acidification response agree very well with pulse-induced stomatal opening seen in intact leaves⁶. It is noteworthy that with both phenomena, responsiveness to a second pulse progressed with nearly identical kinetics, with half-maximal response occurring at a 10-min interval between pulses (Fig. 2 and ref. 6). The similarity indicates that the cellular events underlying the blue light-induced acidification in protoplasts are components of the sensory transduction process mediating the blue light-dependent stomatal movement. This view was further supported by the finding that blue light-induced medium acidification was inhibited, though partially, by the phytohormone abscisic acid (Fig. 3), an important mediator of the closure response of stomata to water stress^{8,14}.

Medium acidification could be a result of cellular H^+ extrusion or of increased respiration rates¹⁵ and concomitant CO_2 evolution. CO_2 in solution equilibrates with HCO_3^- through the reaction $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{HCO}_3^- + \text{H}^+$. Thus, an acidification caused by CO_2 release should be markedly dependent on medium pH (dotted line in Fig. 1*D*). Data presented in Fig. 1*D* show that the blue light-induced acidification was independent of medium pH from 5 to 7, ruling out a significant contribution to the measurements from CO_2 evolution. Further evidence against respiration as a primary cause of the acidification was obtained with measurements of O_2 tension in the suspension medium, by using a Clark-type O_2 microelectrode. The measurements failed to show any detectable change in O_2 levels during blue light-induced acidification (not shown). These data indicate that the observed acidification was primarily a result of H^+ extrusion by guard-cell protoplasts.

In recent work with the electrophysiological technique of patch clamping, *Vicia* guard cell protoplasts clamped at a constant current showed a membrane hyperpolarization in response

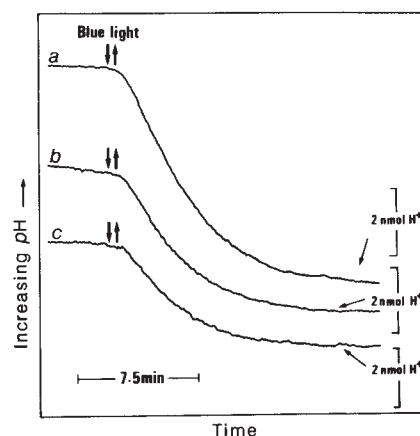


Fig. 3 Effect of abscisic acid on the blue light-induced acidification by guard-cell protoplasts. The extent of acidification induced by a saturating pulse of blue light (trace *a*) was reduced by 40% by an addition of $5 \mu\text{M}$ abscisic acid (trace *b*). A higher concentration of abscisic ($50 \mu\text{M}$) increased the inhibition only slightly (trace *c*). Abscisic acid was added to the protoplast suspension 20 min before the blue light treatment (30 s with $50 \mu\text{mol m}^{-2} \text{s}^{-1}$). Experiments repeated three times on different occasions showed similar results. Other details as in Fig. 1.

to a blue-light pulse under a background of red light¹⁶. Under constant voltage, blue light pulses elicited a transient outward current, with both the hyperpolarization and current measurements¹⁶ showing time courses similar to those seen with the blue light-induced acidification reported here. Maximal rates of H^+ extrusion per plasmalemma unit area, estimated from results of many experiments, ranged between 2 and $6 \text{ pmol } H^+ \text{ cm}^{-2} \text{ s}^{-1}$. These values agree closely with the measured outward current density of $0.88 \mu\text{A cm}^{-2}$ (ref. 16). The temporal homology between acidification and hyperpolarization responses, and the quantitative agreement between the measured current density and H^+ release rate strongly indicate that the blue light-depen-

Table 1 Effects of vanadate, diethylstilboestrol and CCCP on the proton extrusion induced by a blue light pulse and fusicoccin

Inhibitor	Concentration	Rate of proton release (% of control)	
		Blue light pulse	Fusicoccin
Diethylstilboestrol	$50 \mu\text{M}$	0	0
CCCP	$10 \mu\text{M}$	0	0
Vanadate (pH 6.2)	0.16 mM	ND	119
	0.83 mM	113	ND
	1.0 mM	98	ND
	1.67 mM	ND	106
Vanadate (pH 7.0)	2.5 mM	108	ND

Fusicoccin (Farmopant) was applied at $15 \mu\text{M}$ to a 1.2-ml suspension of guard-cell protoplasts kept under continuous red light ($800 \mu\text{mol m}^{-2} \text{s}^{-1}$). Proton extrusion started immediately after addition of fusicoccin, at rates ranging between 9.4 and $12.8 \text{ pmol cm}^{-2} \text{s}^{-1}$ and continued for at least 15 min. Control rates for the blue light-induced proton extrusion (30-s pulse at $50 \mu\text{mol m}^{-2} \text{s}^{-1}$) ranged between 2.4 and $5.6 \text{ pmol cm}^{-2} \text{s}^{-1}$. Rates of proton extrusion in the presence of inhibitors were given as percentage of control values. Rates of H^+ extrusion were calculated as in Fig. 1. CCCP, diethylstilboestrol and vanadate were added to the suspension medium from stock solutions, 30 min before the application of a blue light pulse or the addition of fusicoccin. The vanadate solution at $20 \mu\text{M}$, pH 7.0, inhibited a *Neurospora* plasmalemma ATPase preparation by 90% (C. Borgeson, personal communication). 'Zero' indicates no detectable change in pH baseline levels throughout the time course of the observed response in the presence of inhibitor; ND, no determined. In all cases, the blue light pulse was given on a background of continuous red light.

dent, electrogenic ion pump resolved in the path-clamp experiments¹⁶ is a H⁺ pump. The results support a chemi-osmotic mechanism as the basis of ion uptake by guard cells¹.

Proton pumping in plant cells is usually attributed to the activity of a plasmalemma ATPase¹⁷, with its activity reported to be inhibited by vanadate¹⁸. However, vanadate failed to affect the blue light-induced acidification, even at 2.5 mM (Table 1). Because this finding could imply that blue light-activated H⁺ extrusion is mechanistically different from other H⁺-pump systems previously described¹⁷, we tested the vanadate sensitivity of the fusicoccin-induced H⁺ extrusion, which is known to be mediated by a plasmalemma ATPase¹⁹. The high rate of H⁺ extrusion elicited by fusicoccin in guard-cell protoplasts was also not affected by vanadate (Table 1). Inability of guard-cell protoplasts to take up vanadate could explain the results¹⁸. On the other hand, blue light-induced acidification was abolished by diethylstilboestrol, a putative inhibitor of plasmalemma ATPase²⁰ (Table 1). Carbonyl cyanide-*m*-chlorophenylhydrazone (CCCP) also completely suppressed both the blue light- and fusicoccin-dependent acidification (Table 1). CCCP is a protonophore that dissipates H⁺ gradients; it can also suppress ATPase activity by reducing ATP levels in guard-cell protoplasts²¹. Although the results from the inhibitor experiments described above are in accord with the possible mediation of blue light-induced H⁺ extrusion by a plasmalemma ATPase, conclusive demonstration awaits further investigation.

The acidification response to blue-light pulses has further implications for the role of blue light in the activation of the H⁺ pump. The observed high rates of H⁺ release for minutes after pulse application argues against a direct energetic contribution by blue light for H⁺ extrusion. Furthermore, conservatively assuming a H⁺/ATP ratio of 2 (ref. 17), a calculation of the maximal efficiency of light quanta for H⁺ extrusion, estimated from the initial slope of the data shown in Fig. 1C, indicates that this blue light response is about 15-fold more efficient, based on a unit chlorophyll weight, than the ATP formation from non-cyclic photophosphorylation, measured in *Vicia* guard-cell chloroplasts¹⁰. Thus, the activated, blue light receptor probably functions as a modulator of H⁺-pumping activity, and not as a source of energy for the pump. Further characterization of sensory transduction in guard cells should increase our understanding of the photobiological and bioenergetic properties of the blue light response⁷.

We thank Dr B. B. Jørgensen (Aarhus University, Denmark) for the O₂ microelectrode methodology and equipment. This research was supported by the NSF, the US Department of Energy and the Department of Agriculture. K.S. gratefully acknowledges a McKnight postdoctoral fellowship. M.I. thanks Dr W.R. Briggs for generous encouragement during this work. Publication no. 903 of the Carnegie Institution of Washington, Department of Plant Biology.

Received 24 June; accepted 7 November, 1985.

1. Zeiger, E. A. *Rev. Pl. Physiol.* **34**, 441-475 (1983).
2. Hsiao, T. C., Allaway, W. E. & Evans, L. T. *Pl. Physiol.* **51**, 82-88 (1973).
3. Zeiger, E. & Hepler, P. K. *Science* **196**, 887-889 (1977).
4. Ogawa, T., Ishikawa, H., Shimada, K. & Shibata, K. *Planta* **142**, 61-65 (1978).
5. Schwartz, A. & Zeiger, E. *Planta* **161**, 129-136 (1984).
6. Iino, M., Ogawa, T. & Zeiger, E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8019-8023 (1985).
7. Senger, H. (ed.) *Blue Light Effects in Biological Systems* (Springer, Berlin, 1984).
8. Gepstein, S., Jacobs, M. & Taiz, L. *Pl. Sci. Lett.* **28**, 63-72 (1982).
9. Zeiger, E., Moody, W., Hepler, P. K. & Varela, F. *Nature* **270**, 270-271 (1977).
10. Shimazaki, K. & Zeiger, E. *Pl. Physiol.* **78**, 211-214 (1985).
11. Shimazaki, K., Gotow, K. & Kondo, N. *Pl. Cell Physiol.* **23**, 871-879 (1982).
12. Raschke, K. & Humble, G. D. *Planta* **115**, 47-57 (1973).
13. Neumann, J. & Levine, R. P. *Pl. Physiol.* **47**, 700-704 (1971).
14. Hornberg, C. & Weiler, E. W. *Nature* **310**, 321-324 (1984).
15. Kowallik, W. A. *Rev. Pl. Physiol.* **33**, 51-72 (1982).
16. Assmann, S. M., Simoncini, L. & Schroeder, J. I. *Nature* **318**, 285-287 (1985).
17. Spanswick, R. M. A. *Rev. Pl. Physiol.* **32**, 267-289 (1981).
18. Lew, R. R. & Spanswick, R. M. *Pl. Physiol.* **75**, 1-6 (1984).
19. Marré, E. A. *Rev. Pl. Physiol.* **30**, 273-288 (1979).
20. Balke, N. E. & Hodges, T. K. *Pl. Sci. Lett.* **10**, 319-325 (1977).
21. Shimazaki, K., Gotow, K., Sakaki, T. & Kondo, N. *Pl. Cell Physiol.* **24**, 1049-1056 (1983).
22. Arnon, D. *Pl. Physiol.* **24**, 1-15 (1949).

Magnocellular axons in passage through the median eminence release vasopressin

Megan C. Holmes*, Ferenc A. Antoni*†‡, Greti Aguilera* & Kevin J. Catt*

* Endocrinology and Reproduction Research Branch, National Institute of Child Health and Development, and
† Laboratory of Cell Biology, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892, USA

Vasopressin (arginine vasopressin, AVP) is present in two types of nerve fibres in the median eminence (ME)¹⁻³. First, it is found in nerve terminals that originate in the parvicellular neurones of the hypothalamic paraventricular nucleus (PVN) and abut on the pericapillary space surrounding the fenestrated capillaries of the primary pituitary portal plexus in the external zone (EZ) of the ME. These neurones also synthesize corticotropin-releasing factor (CRF)⁴⁻⁶, which acts synergetically with vasopressin to stimulate release of adrenocorticotropin (ACTH) from the pituitary gland (see ref. 7). Second, vasopressinergic axons of the magnocellular neurosecretory system pass through the internal zone (IZ) of the ME to terminate in the neurohaemal contact zone of the neurohypophysis. The involvement of vasopressinergic magnocellular neurones in the control of ACTH secretion is much debated⁷. Of particular interest in this context is the origin of the vasopressin found in pituitary portal blood^{1,8,9}. Although it has been demonstrated that vasopressin and CRF are present in the same neurosecretory granules of EZ fibres³, parallel determinations of vasopressin and CRF in pituitary portal blood have shown alterations of the concentration of vasopressin without a concomitant change in that of CRF^{8,9}. Such a dissociation suggests that either differential release of vasopressin and CRF can occur from a single population of nerve endings, or there are fibres in the pituitary-stalk ME which release vasopressin but not CRF. Here we present evidence for the latter. Our results indicate that stimuli causing depolarization of the axonal membrane *in vitro* elicit release of vasopressin from nerve fibres in the external and internal zones of the ME.

We have reported that K⁺ ions and veratridine stimulate secretion of CRF and vasopressin from the ME *in vitro* in a dose-dependent manner, and that this process depends on extracellular Ca²⁺ (ref. 10). Immunocytochemical and radio-immunoassay studies have shown that the amount of CRF in the magnocellular fibre system of the ME is at least one order of magnitude lower than that in the EZ parvicellular fibres (see ref. 7). Hence, the release of CRF appears to be an appropriate marker for the secretory activity of the CRF/AVP nerve endings of the EZ; we have used this to analyse the possible compartments of vasopressin release by the ME *in vitro*, after impairing the function of the fibres in the EZ by various experimental manipulations.

First, rats were pretreated with reserpine, which activates the hypothalamo-pituitary-adrenocortical system by inducing hypersecretion of CRF¹¹. As a result, the peptide content of the CRF/AVP terminals in the EZ is dramatically decreased but there is no apparent change in the vasopressin content of the IZ magnocellular fibres^{12,13}. In our hands, pretreatment with reserpine decreased the CRF content of the ME by 64% but had no effect on the amounts of vasopressin in either the ME or the neurointermediate lobe of the pituitary (Fig. 1). The latter result suggests that the quantity of vasopressin in the EZ system is close to that of CRF, that is, small compared with the total ME vasopressin content, and hence minor changes of EZ

‡ Present address: Department of Human Anatomy, University of Oxford, Oxford OX1 2JD, UK.

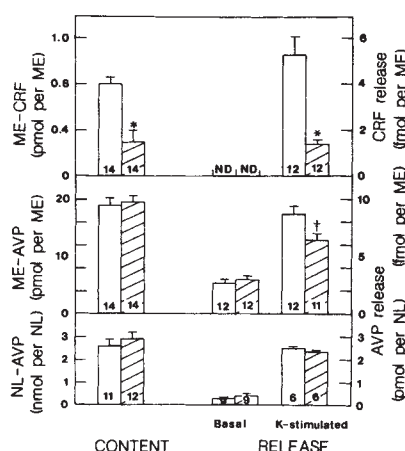


Fig. 1 Effects of pretreatment with reserpine on the content (left) and *in vitro* release (right) of vasopressin and CRF in ME or neurointermediate lobes of the pituitary gland (NL). Tissues were obtained from male Sprague-Dawley albino rats (200–300 g body weight) injected s.c. with either 10 mg per kg reserpine dissolved in dimethyl sulfoxide (DMSO) (hatched columns) or 2 ml per kg of the vehicle alone (open columns) 24 h before killing by decapitation. The columns represent the mean $\pm$ s.e.m., and the numbers of observations are shown at the bottom of each column. The piece of tissue referred to as median eminence also contained variable portions of the pituitary stalk but never any part of the neural lobe of the pituitary gland. The relative amounts of vasopressin released from the two pools of vasopressin in the ME may be estimated as follows: assuming that a proportionate amount of vasopressin is co-released with CRF from a population of parvocellular nerve endings³, the secretion of vasopressin from these fibres should decrease to the same extent as the release of CRF, that is, by 75%. Hence, 75% of AVP released from the CRF/AVP compartment is equivalent to the 44% decrease in the total K^+ -stimulated release of AVP (release in the presence of K^+ minus basal release) provided that reserpine does not affect the release of vasopressin from a second compartment that has no CRF. From this, the amount of vasopressin released by the CRF/AVP compartment in untreated rats can be estimated as 44/75, which is $\sim$ 60% of the total release of vasopressin. The remaining 40% of the vasopressin secreted would originate from the reserpine-insensitive compartment.

Methods. The incubations were performed in 350 μ l of Krebs-Ringer-bicarbonate (KRBG) medium, containing 11.1 mM D-glucose 0.1 mM L-ascorbate and bovine serum albumin (1 g l⁻¹), at 37 °C in an atmosphere of O₂/CO₂ (95:5). After dissection from the brain, the median eminences (6–7 per tube) were preincubated for 30–45 min, after which the medium was carefully removed and changed every 10 min during the period of the experiment for any one of the following: fresh KRBG, KRBG with 48 mM K^+ , from which an appropriate amount of NaCl was omitted to maintain iso-osmotic conditions; KRBG with 10 μ M veratridine (supplied by Dr G. Brown, University of Alabama, Birmingham); KRBG containing 1 mM EGTA and no added calcium; KRBG containing 1 μ M tetrodotoxin. The concentrations of CRF and vasopressin in the incubation medium were determined by specific radioimmunoassays as described elsewhere¹⁰. After the incubation, the tissues were homogenized in 100 μ l of 0.1 M HCl containing 1 mM ascorbate, and after appropriate dilution in 0.05 M sodium phosphate buffer containing 0.1% Triton X-100 (v/v), the peptide content was determined by radioimmunoassay. * $P < 0.01$; † $P < 0.05$ by Student's *t*-test in comparison with the group treated with vehicle alone.

vasopressin content are unlikely to be detected. In parallel with the decrease of CRF content, the K^+ -stimulated release of CRF was reduced by 75% in rats treated with reserpine (Fig. 1). Pretreatment with reserpine also diminished the K^+ -evoked release of vasopressin from the ME, but to a lesser extent (44%) than that of CRF. In contrast, no change of the K^+ -induced release of vasopressin from neurointermediate lobes was found (Fig. 1). These findings are consistent with the suggestion that

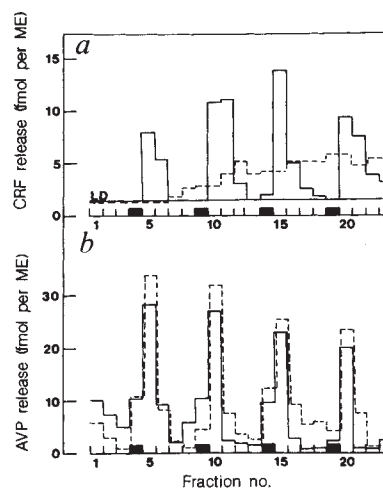


Fig. 2 Potassium-evoked release of CRF (a) and vasopressin (b) by median eminences obtained from sham-operated (solid line) or 7-day PVN-lesioned (broken line) rats. The solid bars indicate the application of KRBG containing 48 mM KCl. The abscissa shows the number of 10-min fractions collected. In a, the continuous line at the bottom of the figure, labelled LD, indicates the limit of detection in the radioimmunoassay. Data are representative of six pairs of similar experiments.

Methods. Lesioning of the PVN was performed in rats anaesthetized with chloral hydrate (4 g per kg body weight) using a specially designed microknife¹⁵. After killing by decapitation on post-operative day 7, the site of the lesion in each brain was examined in serial coronal sections (300–400 μ m thick)¹⁵ prepared during the period of preincubation of the median eminences in individual wells of a 24-well polystyrene tissue-culture plate. Subsequently, the tissues from sham-operated rats or those with complete lesions of the PVN were pooled (6–7 per tube) and incubated as described in Fig. 1 legend. The CRF content of each ME was measured after the incubation as an additional indicator of the completeness of the lesion.

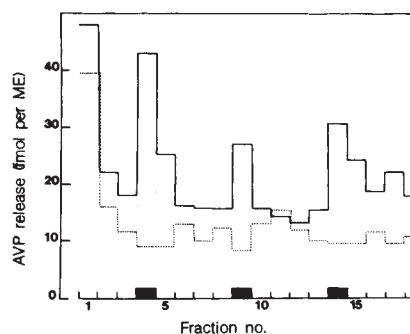


Fig. 3 Potassium-evoked release of vasopressin by median eminences obtained from sham-operated rats (solid line) or from animals subjected to anterolateral transection of the hypothalamus 7 days before the experiment (dotted line). Solid bars indicate the application of KRBG containing 48 mM KCl. The abscissa shows the number of 10-min fractions collected. The data are representative of three pairs of similar experiments. Three ME were incubated per tube as described in Fig 1 legend. Anterolateral cuts were made with a bayonet-shaped knife, in rats anaesthetized with chloral hydrate as described previously¹⁸. Each brain was examined to determine the site of the cut as described in Fig 2 legend.

secretion of vasopressin occurs from two compartments in the ME. One of these compartments also releases CRF and is depleted of peptides by reserpine, whereas the other is not affected by the drug and neither contains nor releases CRF. Neuroanatomical evidence^{1–7,12,13} indicates that the CRF/AVP fibre compartment is present in the EZ, whereas the compartment containing only vasopressin and which is not affected by reserpine is localized in the IZ of the ME.

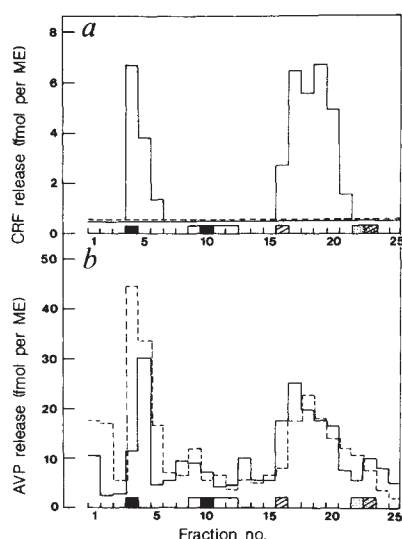


Fig. 4 Release of CRF (a) and vasopressin (b) *in vitro* by median eminences of PVN-lesioned (broken line) and sham-operated (solid line) rats. The abscissa indicates the number of 10-min fractions collected. Solid bars indicate the application of 48 mM KCl; open bars, Ca^{2+} -free medium; hatched bars, $10 \mu\text{M}$ veratridine; stippled bars, $1 \mu\text{M}$ tetrodotoxin. Data are representative of four pairs of similar incubations.

To obtain further evidence for the release of vasopressin from magnocellular axons in the IZ, we examined the secretion of vasopressin and CRF by ME obtained from rats which had been subjected to surgical ablation of the PVN. It has been shown previously that such lesions decrease drastically the CRF content of the ME, as well as the numbers of the nerve fibres that are immunoreactive for CRF and vasopressin in the EZ (see, for example, refs 1, 7). In contrast, the magnocellular neurones of the supraoptic nucleus and the supraoptico-hypophyseal tract remain intact after PVN lesions. In agreement with these findings, we observed a marked decrease in the CRF content of the ME 7 days after hypothalamic surgery: from 848 ± 88 fmol in sham-operated controls to 75 ± 16 fmol in lesioned animals (mean $\pm$ s.e.m., $n = 11$ in both groups). The vasopressin content of the ME was also decreased, from 22.9 ± 2.9 pmol in controls to 12.0 ± 0.8 pmol in lesioned rats (mean $\pm$ s.e.m., $n = 10$ and 11, respectively). These results are consistent with the neuro-anatomical observation that $\sim 40\%$ of the vasopressinergic magnocellular neurones lie in the region destroyed by the PVN lesion¹⁴. There was no release of CRF in response to K^+ in ME obtained from PVN-lesioned rats (Fig. 2). In contrast, K^+ released substantial amounts of vasopressin from the same segments of tissue that showed no CRF response. In three series of experiments, the average amounts of vasopressin released by a 10-min exposure to high- K^+ medium were similar in sham-operated and PVN-lesioned rats (15.0 ± 2.2 and 16.0 ± 3.0 fmol per ME; mean $\pm$ s.e.m., $n = 15$ and 13, respectively). This finding contrasts with the 40% decrease of vasopressin content mentioned above, and suggests that a compensatory increase of releasable vasopressin occurs in the ME after PVN lesions. The nature of this apparently compensatory response is not clear, however it was not inhibited by dexamethasone ($250 \mu\text{g}$ per kg given to PVN-lesioned rats subcutaneously (s.c.) for 7 days; data not shown), making it unlikely that the build-up of vasopressin occurs in surviving cells of the parvocellular CRF/AVP system, which are strongly inhibited by glucocorticoid hormones^{1,15}. Moreover, to compensate for the loss of 90% of the CRF/AVP neurones, the cells spared by the lesion would have to increase their ability to secrete vasopressin by 900% in 7 days; this is also improbable, as in the absence of glucocorticoid feedback following bilateral adrenalectomy, it takes 3–4 weeks for ME vasopressin release to increase by

500%¹⁰. Moreover, PVN-lesioned rats have high concentrations of circulating corticosterone ($\sim 40\%$ of controls)^{16,17}, and circadian rhythms of corticosterone are also apparent¹⁶.

To demonstrate that degenerating magnocellular fibres cannot be the source of the vasopressin release in response to depolarization, ME obtained from rats with anterolateral hypothalamic transections¹⁸ were tested for their ability to release the peptide (these lesions transect all vasopressinergic axons, including those in the supraoptico-hypophyseal tract, that project to the ME). Seven days after the lesion, the vasopressin content of the ME was decreased by $\sim 90\%$ (2 ± 0.6 pmol per ME, $n = 9$ compared with 18 ± 3 , $n = 9$ in controls; mean $\pm$ s.e.m.) and there was no detectable release of vasopressin in response to a high K^+ concentration (Fig. 3).

The release of AVP from the ME of sham-operated or PVN-lesioned rats had similar properties (Fig. 4): the release evoked by K^+ was calcium-dependent, and that elicited by veratridine was blocked by tetrodotoxin. The activity of the cytoplasmic enzyme lactate dehydrogenase¹⁹ (EC 1.1.1.27) was not above background (0.001 extinction units per 120 s) in high- K^+ medium bathing ME from PVN-lesioned rats, but on addition of 0.1% Triton X-100, enzyme activity increased 14-fold, indicating leakage of the enzyme from the cytoplasm. Taken together, these data suggest that the release of vasopressin from the ME of PVN-lesioned rats occurs from magnocellular fibres of the supraoptico-hypophyseal tract in the IZ of the ME, via a physiologically relevant exocytotic process.

The magnocellular axons in passage through the IZ are varicose unmyelinated fibres with distinct swellings, called Herring bodies, in which mitochondria and neurosecretory vesicles containing immunoreactive vasopressin are abundant^{20,21}. It appears that the cellular machinery necessary for neurotransmitter release is present along the entire length of the axon, albeit at lower density than at nerve terminals (compare with refs 22, 23), and it is well documented that most of the vasopressin found in the ME is in the biologically active form (see refs 24, 25). Hence, a small but significant fraction of the peptide present in IZ fibres may be released in response to depolarizing stimuli, perhaps preferentially at Herring bodies. The released peptide could gain access to the pituitary portal circulation via the extensive network of fenestrated primary portal capillaries which penetrates the IZ^{20,21} and is not separated by any morphologically evident diffusion barrier from the fibres in passage^{21,22}. Because the amount of vasopressin in the IZ fibre system is about 10–20-fold higher than in the EZ, release from the IZ neurites may account for a substantial proportion ($\sim 40\%$; see Fig. 1 legend) of the vasopressin in pituitary portal blood.

Although we have no direct evidence that the release process described above operates *in vivo*, a considerable amount of functional data favours the existence of two independent sources of ME vasopressin secretion. It has been noted that the concentrations of vasopressin and CRF may vary independently in pituitary portal blood (see refs 8, 9). Furthermore, PVN lesions prevent the increase of plasma ACTH after adrenalectomy, a manoeuvre that stimulates parvocellular CRF/AVP neurones, but not, or only slightly, the magnocellular system^{1,25,26}. In contrast, PVN lesions merely attenuate the pituitary-adrenocortical response to stressful stimuli such as ether-venesection and ether-laparotomy^{16,17}, which are known to activate the parvocellular CRF/AVP system⁷ as well as the magnocellular vasopressinergic neurones^{8,27}. Anterolateral hypothalamic transections that destroy all vasopressinergic fibres (parvocellular as well as magnocellular) reaching the ME completely block the ACTH response to adrenalectomy or the stressful stimuli mentioned above^{7,17}.

Intriguingly, these functional findings are paralleled by studies on the release of vasopressin from the ME *in vitro*, in which lesions of the PVN abolished the enhancement by bilateral adrenalectomy of the depolarization-induced release of

vasopressin²⁶. Here we have reported that lesioning of the PVN in otherwise intact rats did not diminish substantially the release of vasopressin from the ME, whereas anterolateral cuts abolished the vasopressin response to depolarization.

Thus, our data suggest that the original, and now largely rejected, proposal that magnocellular neurones contribute to the regulation of anterior pituitary hormone secretion by releasing neurohormones in the ME²⁸, should be reconsidered with respect to the control of ACTH secretion.

We thank Drs M. H. Whitnall, E. Mezey and H. Gainer for communicating their findings to us before publication.

5. Sawchenko, P. E., Swanson, L. W. & Vale, W. W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1883-1887 (1984).
6. Tramu, G., Croix, G. & Pillez, A. *Neuroendocrinology* **37**, 467-469 (1983).
7. Makara, G. B., Antoni, F. A., Stark, E. & Kerteszi, M. *Perspectives in Neuroendocrinology* Vol. 3 (eds Müller, E. E. & McLeod, R. M.) 71-119 (Elsevier, Amsterdam, 1984).
8. Gibbs, D. M. *Fedn Proc.* **44**, 203-206 (1985).
9. Plotsky, P. M. *Fedn Proc.* **44**, 207-213 (1985).
10. Holmes, M. C., Antoni, F. A., Catt, K. J. & Aguilera, G. *Neuroendocrinology* (in the press).
11. Bhattacharya, A. N. & Marks, B. H. *J. Pharmac. exp. Ther.* **165**, 108-116 (1969).
12. Seybold, V., Elde, R. & Hokfelt, T. *Endocrinology* **108**, 1803-1809 (1981).
13. Bugnon, C. *et al. J. Steroid Biochem.* **20**, 183-195 (1984).
14. Rhodes, C. H., Morrell, J. L. & Pfaff, D. W. *J. comp. Neurol.* **198**, 45-64 (1984).
15. Antoni, F. A., Holmes, M. C. & Kiss, J. Z. *Endocrinology* **117**, 1293-1299 (1985).
16. Ixart, G. *et al. Neuroendocrinology* **35**, 270-276 (1982).
17. Makara, G. B. *Adv. physiol. Sci.* **14**, 31-44 (1981).
18. Antoni, F. A., Makara, G. B. & Rappay, Gy. *J. Endocr.* **91**, 415-423 (1981).
19. Marchbanks, R. M. *Biochem. J.* **104**, 148-157 (1967).
20. Seitz, H. M. Z. *Zellforsch. Mikrosk. Anat.* **67**, 351-366 (1965).
21. Page, R. B. & Dovey-Hartman, J. J. *comp. Neurol.* **226**, 274-288 (1984).
22. Grafstein, B. & Forman, D. S. *Physiol. Rev.* **60**, 1167-1283 (1980).
23. Reichardt, L. F. & Kelly, R. B. A. *Rev. Biochem.* **52**, 871-926 (1983).
24. Gillies, G. & Lowry, P. J. *Frontiers in Neuroendocrinology* Vol. 7 (eds Ganong, W. F. & Martini, L.) 45-75 (Raven, New York, 1981).
25. Russell, J. T., Brownstein, M. J. & Gainer, H. *Brain Res.* **201**, 227-232 (1980).
26. Knepel, W., Nütto, D., Meyer, D. K. & Vlaskovska, M. *Neurosci. Lett.* **48**, 321-326 (1984).
27. Andersson, E. *Science* **152**, 379-380 (1966).
28. Green, J. D. & Harris, G. W. *J. Endocr.* **5**, 136-145 (1947).

Received 27 August; accepted 21 November 1985.

1. Zimmerman, E. A. *et al. Ann. N.Y. Acad. Sci.* **297**, 405-417 (1977).
2. Burlet, A., Chateau-Chapler, M. & Dreyfuss, F. *Neuroendocrinology of Vasopressin, Corticotiberin and Opiomelanocortins* (eds Baertschi, A. J. & Dreyfuss, J. J.) 95-106 (Academic, London, 1982).
3. Whitnall, M. H., Mezey, E. & Gainer, H. *Nature* **317**, 248-250 (1985).
4. Kiss, J. Z., Mezey, E. & Skirboll, L. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1854-1858 (1984).

Excitatory amino acids inhibit stimulation of phosphatidylinositol metabolism by aminergic agonists in hippocampus

Michel Baudry, Jim Evans & Gary Lynch

Department of Psychobiology and Center for the Neurobiology of Learning and Memory, University of California, Irvine, California 92717, USA

Since the initial observations in the 1950s^{1,2} a large number of neurotransmitters and hormones have been shown to influence phosphatidylinositol (PI) metabolism in brain and peripheral ganglia (see ref. 3 for review). This has led to the suggestion that PI is part of an intracellular second messenger system for some types of diffusible chemical factors^{4,5}. Consistent with this are recent reports that one of the products of PI turnover (diacylglycerol) stimulates the Ca-dependent phospholipid-dependent protein kinase (kinase C) while a second (inositol trisphosphate) causes the release of calcium from intracellular stores^{6,7}. Thus it is possible that at least some brain neurotransmitters utilize the PI system to produce functional effects that are in addition to and which outlast the very brief physiological responses they elicit. Although it had been anticipated that another class of receptors might inhibit receptor-mediated stimulation of PI breakdown, no clear examples of such effects have been described. We now report that acidic amino acids, which are thought to be excitatory neurotransmitters at the majority of brain synapses⁸, strongly inhibit the stimulation of PI metabolism elicited by carbachol, histamine, or by potassium-induced depolarization, without changing the response to noradrenaline. As well as indicating a novel function for the excitatory amino acids, these results suggest that the central nervous system possesses cell-cell interactions of a previously unsuspected type.

The experiments were performed on transverse rat hippocampal slices prepared as described previously⁹. Following preincubation for 1 h at 33 °C in Krebs-bicarbonate medium under a constant stream of O₂/CO₂ (95:5), the slices were further preincubated for 1 h at 33 °C in the presence of ³H-*myo*-inositol (NEN, 15-20 Ci mmol⁻¹). The accumulation of ³H-inositol phosphate derivatives (³H-InsP) induced by various agonists was then determined in the presence of 5 mM lithium chloride as described by Brown *et al.*¹⁰ (see Fig. 1). The elution profile of the ³H-inositol phosphates obtained by stepwise addition of formate solutions of increasing strength indicated that more than 90% of these derivatives corresponded to *myo*-inositol-1-phosphate, a result in good agreement with that of Brown *et*

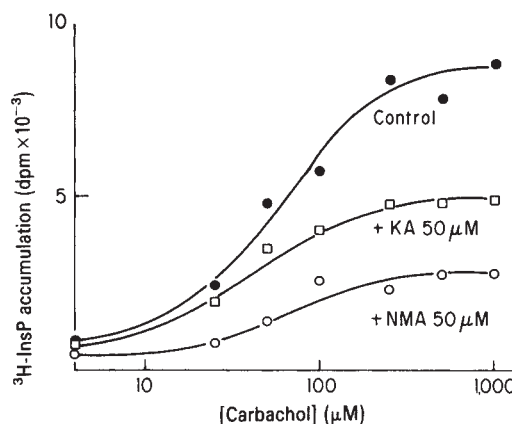


Fig. 1 Effects of *N*-methylaspartic acid (NMA) and kainic acid (KA) on basal and carbachol-stimulated accumulation of ³H-InsP in rat hippocampal slices. Transverse hippocampal slices (0.4 mm thick) were cut with a McIlwain tissue chopper and preincubated for 60 min at 33 °C in Krebs-Ringer bicarbonate medium containing (final concentration in mM): NaCl 124, KCl 3.33, Mg₂SO₄ 2.42, CaCl₂ 3.1, KH₂PO₄ 1.25, NaHCO₃ 25, D-glucose 10, urea 2 and ascorbic acid 2, under a constant stream of O₂/CO₂ (95:5). After three washes with freshly oxygenated medium, slices were further incubated for 60 min at 33 °C in the same medium containing ³H-*myo*-inositol (NEN, 15.8 Ci mmol⁻¹, 5 μCi ml⁻¹). Slices were washed again three times with freshly oxygenated medium and were distributed in incubation vials (two or three slices per vial) and incubated for 10 min at 33 °C in the presence of 5 mM LiCl and various concentrations of carbachol and excitatory amino acids (final volume 0.25 ml). At the end of the incubation 940 μl of a mixture of chloroform/methanol (1:2) were added, slices were sonicated briefly and 310 μl of chloroform and 310 μl of H₂O were added to separate the phases. Aliquots of the upper aqueous phase (0.75 ml) were transferred to a column containing ~175 mg of an anion-exchange resin (Bio-Rad AG 1X-8, 100-200 mesh, formate form). The columns were washed with 10 ml of H₂O to remove ³H-*myo*-inositol. Labelled ³H-inositol phosphate derivatives (³H-InsP) were eluted with 5 ml of 200 mM ammonium formate/100 mM formic acid and counted in a liquid scintillation counter, and are expressed as d.p.m. of ³H-InsP accumulated and represent means of 4-6 experiments with s.e.m. less than 10% of the means.

*al.*¹⁰. Figure 1 indicates that two different excitatory amino acids, *N*-methylaspartic acid (NMA) and kainic acid (KA) had a small although not statistically significant inhibitory effect on basal accumulation of ³H-InsP. As previously reported by various groups^{10,11}, carbachol caused a large and concentration-dependent stimulation of ³H-InsP accumulation. Both NMA (50 μM)

and kainic acid (50 μ M) markedly inhibited this latter effect, apparently by decreasing the maximal response to carbachol without significantly changing its apparent EC_{50} (50% effective concentration) (Fig. 1); this inhibition, as well as that produced by two other excitatory amino acids, was found to be concentration-dependent (Fig. 2; NMA, IC_{50} (50% inhibitory concentration) of 20 μ M; D, L-homocysteic acid, IC_{50} of 40 μ M; kainic acid, IC_{50} of 50 μ M; and L-glutamic acid, IC_{50} of 1.5 mM). The inhibitory action of NMA was blocked by two specific antagonists of the NMA receptors, D, L-aminophosphonovaleric acid (APV) and D- α -amino adipate (D-AA) (80 $\pm$ 5% inhibition by 400 μ M APV, 50 $\pm$ 5% inhibition by 2 mM D-AA), while the response to kainic acid was not changed by these compounds. NMA action (but not KA) was also antagonized by ketamine, again suggesting that NMA acts by stimulating a physiological receptor, since it has been shown that dissociative anaesthetics act as NMA antagonists¹². The effect of the excitatory amino acids was not due simply to the depolarization they produce as potassium-induced depolarization increased 3 H-InsP accumulation and left the carbachol response unchanged (data not shown). Calcium did not appear to be involved since the amino acids blocked the stimulation of PI breakdown elicited by carbachol in calcium-free medium containing 0.1 mM (Table 1) or 0.5 mM EGTA (data not shown). Note that Kendall and Nahorski¹³ found the PI response to carbachol to be markedly reduced in tissue sections preincubated in the absence of calcium while, as shown in Table 1, we did not obtain this result in brain slices preincubated in normal media and then tested without calcium. Preincubation presumably results in extremely low levels of extracellular calcium as well as changes in the intracellular distribution of the cation—thus stimulation of the PI system may have an absolute requirement for calcium. But the occurrence of typical cholinergic and amino-acid responses in calcium-free medium supplemented with high concentrations of EGTA (Table 1) indicates that the effects reported here are not due to physiological fluxes of calcium. In contrast, removal of sodium from the incubation medium reduced carbachol-induced 3 H-InsP accumulation and virtually eliminated the inhibitory effects of the amino acids (Table 1); moreover, high concentrations of veratridine, which result in the opening of sodium channels¹⁴, also reduced the stimulation by carbachol of PI metabolism (60% reduction at 100 μ g ml⁻¹ veratridine). To assess the specificity of the inhibition produced by excitatory amino acids, we tested NMA with non-cholinergic agents known to stimulate PI breakdown (Table 2). While 100 μ M NMA

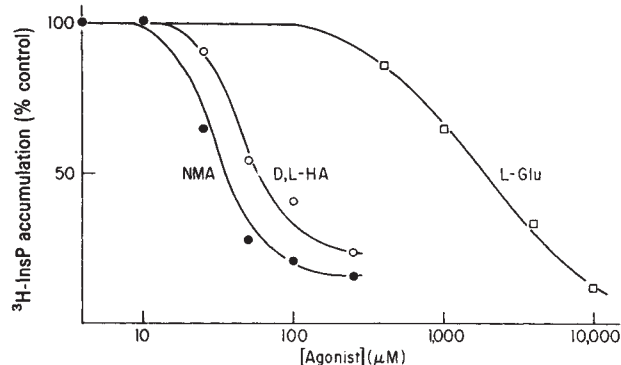


Fig. 2 Effects of various excitatory amino acids on the carbachol-stimulated accumulation of 3 H-InsP in rat hippocampal slices. The effects of various concentrations of different excitatory amino acids—N-methylaspartic acid (NMA), D,L-homocysteic acid (D, L-HA) and L-glutamic acid (L-Glu)—on the carbachol-stimulated accumulation of 3 H-InsP in rat hippocampal slices were determined as described in Fig. 1 legend. Results are expressed as a percentage of the accumulation elicited by carbachol (0.5 mM) in the presence of the indicated concentrations of excitatory amino acids and are means of 4 experiments, with s.e.m. representing less than 10% of the means.

almost totally blocked the stimulation induced by K^+ (50 mM) or histamine (100 μ M) it did not detectably alter the response to noradrenaline (100 μ M).

The present results indicate clearly that activation of excitatory amino-acid receptors suppresses receptor-activated phosphoinositol breakdown. This is evidenced by the fact that the apparent IC_{50} values of various excitatory amino acids in inhibiting the carbachol response are identical to their apparent EC_{50} values for stimulation of receptor-activated sodium fluxes measured under the same conditions¹⁵. Moreover, selective antagonists of certain classes of glutamate receptors¹⁶ block the negative actions of some agonists. There are several routes through which the suppressive effects of the amino acids might be mediated. First, these compounds interact with receptors to open membrane channels for various ion species that might then act on the stimulation of PI metabolism via other receptors. Extracellular calcium is not likely to be involved since both carbachol and the amino acids were effective in calcium-free medium. However, certain of our results point to a build-up of intracellular sodium as a triggering mechanism. As mentioned above, there is a striking correspondence between the concentrations of different amino acids needed for suppression of the response of the PI system to carbachol and those required to promote sodium fluxes; moreover, removal of sodium ions eliminates the effect of the amino acids. While potassium-induced depolarization, which would be expected to cause substantial fluxes of several ion species including sodium, stimulates PI metabolism and does not detectably influence the carbachol-elicited response, the many effects produced by release of neurotransmitters and modulators make the interpretation of

Table 1 Effects of NMA and KA on carbachol stimulation of PI turnover in Na^+ - or Ca^{2+} -free medium

	Basal	Carbachol (0.5 mM)	Carbachol + NMA (50 μ M)	Carbachol + KA (50 μ M)
Normal medium (6)				
3 H-InsP (d.p.m.)	850 $\pm$ 80	8,280 $\pm$ 485	2,355 $\pm$ 600	3,980 $\pm$ 210
% Basal		974 $\pm$ 57	277 $\pm$ 71	468 $\pm$ 25
Na^+ -free medium (4)				
3 H-InsP	2,123 $\pm$ 175	4,931 $\pm$ 302	4,595 $\pm$ 430	4,838 $\pm$ 276
% Basal		232 $\pm$ 14	216 $\pm$ 23	272 $\pm$ 16
Ca^{2+} -free medium (4)				
3 H-InsP	670 $\pm$ 134	9,045 $\pm$ 905	2,438 $\pm$ 138	2,648 $\pm$ 440
% Basal		1,350 $\pm$ 135	364 $\pm$ 21	395 $\pm$ 66

Results are expressed as d.p.m. 3 H-InsP accumulated in the presence of various effectors as well as in percentage of basal values; they represent means $\pm$ s.e.m. of the indicated number of experiments (in parentheses). Hippocampal slices (0.4 mm thick) were preincubated for 60 min at 33 $^{\circ}$ C in normal Krebs-Ringer bicarbonate buffer under a constant stream of O_2/CO_2 (95:5). After washing with fresh medium they were further incubated for 60 min at 33 $^{\circ}$ C in the same medium in the presence of 3 H-myo-inositol (5 μ Ci ml⁻¹). They were then washed with either normal Krebs-Ringer medium (normal medium) or with a medium in which NaCl and NaHCO₃ were substituted with Tris (the pH being adjusted with HCl) (Na^+ -free medium) or with a medium in which CaCl₂ was substituted with MgCl₂ and supplemented with 0.1 mM EGTA (Ca^{2+} -free medium). Slices were then incubated for 10 min at 33 $^{\circ}$ C in the presence of 5 mM LiCl with or without carbachol (0.5 mM), NMA (50 μ M), KA (50 μ M) and combinations of these agonists. 3 H-InsP accumulation was determined as described in Fig. 1 legend.

Table 2 Effect of NMA on 2 H-InsP accumulation elicited by potassium, noradrenaline and histamine in rat hippocampal slices

Agonists	Control	+ NMA (100 μ M)
None	1,253 $\pm$ 110	800 $\pm$ 70
K^+ (50 mM)	4,441 $\pm$ 590	1,747 $\pm$ 225*
NA (100 μ M)	2,947 $\pm$ 265	2,647 $\pm$ 224
Histamine (100 μ M)	2,252 $\pm$ 379	1,286 $\pm$ 125*

Details as for Table 1, except that slices were incubated for 30 min at 33 $^{\circ}$ C in the presence of the various effectors. Results are expressed as d.p.m. 3 H-InsP accumulated and are means $\pm$ s.e.m. of 6–10 experiments.

* $P < 0.01$ (Student's t -test).

the results obtained with this procedure quite difficult. Given the links between intracellular sodium and calcium¹⁷, it is conceivable that the large sodium fluxes elicited by excitatory amino acids alter intracellular levels of calcium which could then be responsible for the suppressive effects on PI metabolism.

Receptor-receptor interactions offer a second means through which the excitatory amino acids might block carbachol-induced ³H-InsP accumulation. There is some evidence that receptor-mediated stimulation of PI turnover involves an interaction with a GTP-binding protein in a manner analogous to the coupling of the adrenergic receptor with adenylate cyclase⁶. Processes such as these might prove to be susceptible to lateral effects set up in the membranes by the binding of ligands to neighbouring non-cholinergic receptors. However, the absence of an NMA effect on stimulation of PI metabolism by noradrenaline would argue against such a mechanism unless the adrenergic and excitatory amino-acid receptors were located in different parts of the cells and/or in different types of cells. There is indeed some evidence for the existence of different pools of inositol phospholipids associated with different receptors and possibly at different cellular locations¹³.

Finally, the excitatory amino-acid receptors could be associated with second messenger systems of their own that, among other things, inhibit the hydrolysis of phosphatidylinositol. Recent studies have shown that phorbol esters inhibit carbachol-induced stimulation of PI metabolism, possibly by activating protein kinase C, suggesting the existence of a negative feedback between kinase C activation and PI metabolism¹⁸.

Whatever mechanism is involved, the results reported here indicate that the large number of transmitter or hormone receptors positively linked to PI metabolism are to some degree balanced by other receptors that suppress the activation of this second messenger system. This offers the potential for interaction between chemical messengers at the level of intracellular transducer systems and strengthens the idea that neurones integrate chemical as well as electrical signals.

This research was supported by grant BNS 8112156 to M.B. and AFOSR 82-0116 B to G.L. We thank Jackie Porter for secretarial assistance.

Received 24 July; accepted 8 November 1985.

- Hokin, L. E. & Hokin, M. R. *Biochim. biophys. Acta* **16**, 229; **18**, 102 (1955).
- Larrabee, M. G. *J. Neurochem.* **15**, 803-808 (1968).
- Michell, R. H. *Biochim. biophys. Acta* **415**, 81-147 (1975).
- Michell, R. H. *Neurosci. Res. Prog. Bull.* **20**, 338-350 (1981).
- Downes, C. P. *Trends Neurosci.* **6**, 313-316 (1983).
- Berridge, M. J. & Irvine, R. F. *Nature* **312**, 315-321 (1984).
- Nishizuka, Y. *Science* **225**, 1365-1370 (1984).
- Roberts, P. J., Storm-Mathisen, J. Jr & Johnston, G. (eds) *Glutamate as a Transmitter* (Wiley, New York, 1980).
- Baudry, M., Siman, R., Smith, E. K. & Lynch, G. *Eur. J. Pharmac.* **90**, 161-168 (1983).
- Brown, E., Kendall, D. A. & Nahorski, S. R. *J. Neurochem.* **42**, 1379-1387 (1984).
- Janowsky, A., Labarca, R. & Paul, S. M. *Life Sci.* **35**, 1953-1961 (1984).
- Thomson, A., West, D. C. & Lodge, D. *Nature* **313**, 479-481 (1985).
- Kendall, D. A. & Nahorski, S. R. *J. Neurochem.* **42**, 1388-1394 (1984).
- Narahashi, T. *Physiol. Rev.* **54**, 813-889 (1974).
- Baudry, M., Kramer, K., Fagni, L., Recasens, M. & Lynch, G. *Molec. Pharmac.* **24**, 222-228 (1983).
- Foster, A. C. & Fagg, G. E. *Brain Res. Rev.* **7**, 103-164 (1984).
- Erulkar, S. D. & Fine, A. *Rev. Neurosci.* **4**, 179-323 (1979).
- Labarca, R., Janowsky, A., Patel, J. & Paul, S. M. *Biochem. biophys. Res. Commun.* **123**, 703-709 (1984).

Expression of a translocated *c-abl* gene in hybrids of mouse fibroblasts and chronic myelogenous leukaemia cells

Danuta Kozbor*, Agata Giallongo*†, Michael E. Sierzega*, James B. Konopka‡, Owen N. Witte‡, Louise C. Showe* & Carlo M. Croce*

* Wistar Institute of Anatomy and Biology, 36th Street at Spruce, Philadelphia, Pennsylvania 19104, USA

‡ Department of Microbiology and Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California 90024, USA

Chronic myelogenous leukaemia (CML) is a clonal disease arising from malignant transformation of pluripotent hematopoietic stem cells. In most cases, it is characterized by the presence of the Philadelphia (Ph¹) chromosome (22q-) which results from a reciprocal translocation between chromosomes 9 and 22 (refs 1-3). In this translocation, the human homologue of the Abelson virus oncogene, *c-abl*, normally on chromosome 9, is moved to chromosome 22, while *c-sis*, the cellular homologue of the simian sarcoma virus oncogene, is moved from chromosome 22 to chromosome 9 (refs 4-6). CML cells carrying the t(9; 22) chromosomal translocation are known to produce an 8-kilobase (kb) *c-abl* transcript in addition to the normal 6- and 7-kb transcripts⁶⁻⁹ and to express the normal p145 *abl* protein and a p210 *c-abl* protein possessing a tyrosine kinase activity not detected in the p145 species¹⁰⁻¹². Results of our analyses using somatic cell hybrids between a mouse fibroblast line and two human CML-derived cell lines which carry the Ph¹ chromosome and are phenotypically identical to the fibroblast parent indicate that only the hybrid cells containing Ph¹ chromosome express both the 8-kb *c-abl* RNA and the p210 protein. Thus, expression of the altered *c-abl* transcripts and protein depends on the presence of the Ph¹ chromosome and is not myeloid-specific.

To establish whether the 8.0-kb *c-abl* aberrant transcript and the p210 protein are derived solely from the translocated *c-abl* oncogene on the Ph¹ chromosome, we constructed somatic cell hybrids between LM-TK⁻ mouse fibroblasts¹³ and K562 (ref. 14) or Bv173 (ref. 15) cells and followed the segregation of the Ph¹ chromosome and the normal chromosome 9 from these hybrids. Because the hybrids are phenotypically indistinguishable from the rodent parental cells, it is also possible to determine whether the expression of the altered *c-abl* transcript requires a myeloid cellular background.

The DNA of both human and mouse parents and of hybrid clones containing different complements of human chromosomes were screened by Southern blot analysis using *c-sis*, constant-region light-chain (*C_λ*) and *v-abl* probes to select hybrids containing either the 9q+ or the 22q- (Ph¹) chromosome (Fig. 1). The *v-abl* probe used was the 1.6-kb *Bgl*II/*Bgl*II fragment from 2.5 to 4.1 kb on the Abelson oncogene map¹⁶. The *C_λ* probe was a genomic clone (Chr22λ5) of the *C_λ* gene in λWes (ref. 17). This clone contains an 8.5-kb *Eco*RI fragment that includes *Ke⁻Oz⁻* and *Ke⁻Oz⁺* (refs 17, 18). The *c-sis* probe was constructed by subcloning a 1.7-kb *Bam*HI restriction fragment into the plasmid pBR322; this fragment comprises the 3' exon of the human *c-sis* homologue⁵. With these probes, three hybrid clones (clone 260-3-1-3 from the Bv173 × LM-TK⁻ hybrids and clones KL-17-9-1 and KL-17-9-3 from the K562 × LM-TK⁻ hybrids) were identified which contained the Ph¹ chromosome. All three hybrids (Fig. 1a, b, lanes 4-6) were positive for human *c-abl* (17.0 kb), *C_λ* (17.0, 11.2 and 10 kb) and mouse *c-abl* (7.7 and 3.7 kb) sequences. Only 10-20% of the cells in the KL-17-9-3 clone contained the human *c-abl* oncogene, indicating that the 17.0-kb *c-abl* signal in this clone is very weak (Fig. 1a, lane 5). The human *C_λ* probe did not cross-hybridize with mouse DNA. The variable-region light-chain (*V_λ*), *C_λ* and *c-abl* genes were amplified five- to ninefold in K562 cells¹⁹⁻²¹. The *V_λ*, *C_λ* and *c-abl* sequences were on the same aberrant chromosome derived from the Ph¹ chromosome of K562 cells²⁰. DNA of hybrids KL-17-9-1 and KL-17-9-3 (Fig. 1c, lanes 4, 5) hybridized with both human (19 kb) and

† Present address: Istituto di Biologia dello Sviluppo, Via Archirafi 20, 90100 Palermo, Italy.

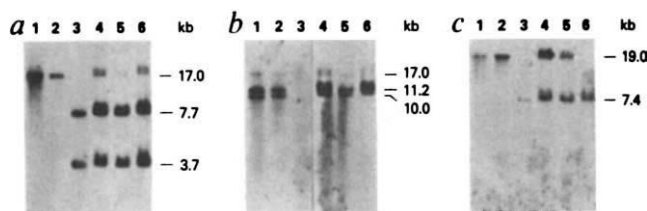


Fig. 1 Southern blot analysis of *Hind*III-digested cellular DNA using *v-abl* (a), *C_λ* (b) and *c-sis* (c) probes. Lane 1, K562 DNA; lane 2, Bv173 DNA; lane 3, LM-TK⁻ DNA; lanes 4 and 5, hybrids KL-17-9-1 and KL-17-9-3 DNA, respectively; lane 6, hybrid 260-3-1-3 DNA. 10 µg of DNA was loaded in lanes 1-3 and 20 µg in lanes 4-6, exposure time was 42 h. Number on the right of each panel indicate size markers.

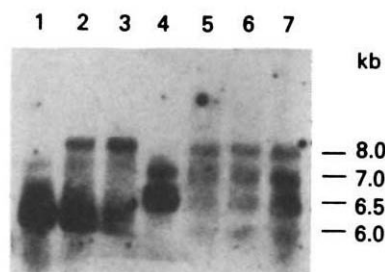


Fig. 2 *c-abl*-related RNA transcripts in PAF (lane 1), K562 (lane 2), Bv173 (lane 3), LM-TK⁻ (lane 4) cells and Ph¹-positive hybrids of K562 and Bv173 with LM-TK⁻ (lanes 5 and 6, hybrids KL-17-9-1 and KL-17-9-3, respectively; lane 7, hybrid 260-3-1-3). Cytoplasmic RNA was extracted by the caesium chloride method²². Total RNA was enriched for poly(A)-containing RNA by one cycle of oligo(dT)-cellulose affinity chromatography³. RNA was denatured in 1 M glyoxal in 10 mM sodium phosphate buffer, pH 7.0, at 50 °C for 1 h, electrophoresed in 1% agarose gel, transferred to nitrocellulose, prehybridized and hybridized to 0.2 µg of the *v-abl* probe according to the method of Thomas²⁴. 10 µg of RNA was loaded in lanes 1-4 and 20 µg in lanes 5-7. Exposure time was 42 h.

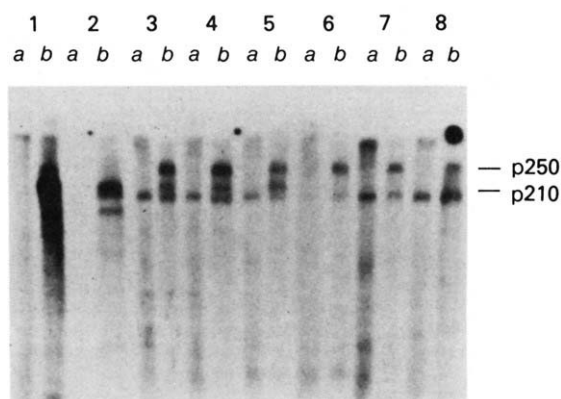


Fig. 3 *In vitro* phosphorylation of altered p210 *c-abl*. For *in vitro* phosphorylation assays, cell extracts were immunoprecipitated with 5 µl of normal rabbit serum (a) or pEX-5 antiserum (b). Lane 1, K562; lane 2, Bv173; lane 3, KL-17-9-1; lane 4, KL-17-9-3; lane 5, 260-3-1-3; lane 6, Ph¹-negative hybrid 260-3-1-3; lane 7, LM-TK⁻; lane 8, IT-22. Immunoprecipitates were washed and incubated in 20 mM PIPES (pH 7.0) containing 10 mM MnCl₂, 10 µM NaF and 5 µCi [γ -³²P]ATP for 5 min at 30 °C. Samples were washed and analysed by SDS-polyacrylamide gel electrophoresis and autoradiography for 2 days.

mouse (7.4 kb) *c-sis*, indicating that these hybrids contain either both the normal chromosome 22 (*C_λ*- and *c-sis*-positive) and normal chromosome 9 (*c-abl*-positive) alternatively the 9q+ chromosome (*c-sis*-positive) and the Ph¹ chromosome. Isoenzyme analysis for adenylate kinase (AK)-1 and -3, chromosome 9 markers (AK-3 is a marker for region 9p13-p24; AK-1 is a marker for region 9q33-qter)²², revealed that neither the normal chromosome 9 nor the 9q+ chromosome is present in the hybrids. Only hybrid 260-3-1-3 (Fig. 1c, lane 6), which is negative for *c-sis* but positive for *C_λ* and *c-abl*, contains only the Ph¹ chromosome, while the other two contain the the Ph¹ chromosome and the normal chromosome 22 (Table 1).

To analyse expression of the altered 8-kb *c-abl* transcript further, as well as the p210 protein with its associated tyrosine kinase activity, we performed Northern blot analysis of poly(A)-containing RNA extracted from parental and hybrid cells (Fig. 2). Two major transcripts of 6 and 7 kb were detected in simian virus 40-transformed Ph¹-negative human fibroblast (PAF) control cells (lane 1). The levels of the normal transcripts in K562 and the PAF control cells were equivalent, whereas in Bv173 cells they were three- to fivefold less, as estimated by densitometry. A faint band of 6.5 kb was also observed in RNA from PAF cells. In addition to the normal transcripts, both K562 (lane 2) and Bv173 (lane 3) expressed the 8-kb aberrant transcript. Two major bands of 6.5 and 7 kb were detected in the mouse LM-TK⁻ RNA (lane 4), while the K562 and Bv173 hybrids contained the human as well as the mouse *c-abl* transcripts. The 8-kb transcript was also present in all three hybrids containing the Ph¹ chromosome. Hybrids that were negative for the Ph¹ chromosome but which contained either the normal chromosome 22 or 9q+ did not express the 8-kb transcript (data not shown). The expression of the 8-kb transcripts does not require the presence of any other human chromosome, because the only common denominator of all the hybrids expressing the 8-kb transcript was the presence of the Ph¹ chromosome or the Ph¹ chromosome derivative in K562 hybrids. However, we were unable to isolate hybrids containing the normal chromosome 9, as this chromosome is preferentially lost from mouse and human hybrids²³. In addition, little or none of the normal 6-kb *c-abl* transcript was detected in these hybrids, suggesting that these transcripts initiate only from the normal chromosome 9 and not from the *c-abl* locus involved in the translocation.

To correlate the presence of the p210 protein with the 8-kb *abl* RNA in Ph¹ chromosome-positive cells, we assayed hybrid cell extracts for *in vitro* kinase activity. Figure 3b shows the electrophoretic profile of cell extracts immunoprecipitated with the *abl*-specific antiserum pEX-5 and then incubated with [γ -³²P]ATP in conditions allowing autophosphorylation. The pEX-5 antiserum was obtained from a rabbit immunized with a *v-abl* polypeptide prepared in a bacterial expression system and has been shown to precipitate the phosphorylated p210 protein in K562 cells¹⁰⁻¹². The p210 protein possessing phosphokinase activity was present in Ph¹-positive K562 (Fig. 3, lane 1) and Bv173 (lane 2) cells, whereas the normal p145 protein was present at equivalent levels in K562 and Ph¹-negative control HL60 cells, but was not detected in Bv173 cells in the same conditions¹². The latter finding is consistent with the lower levels of normal *c-abl* RNA observed in the Bv173 cells. In all hybrids and in parental LM-TK⁻ mouse cells, an additional protein of relative molecular mass 250,000 (p250) was detected. This protein is also present in another mouse fibroblast cell line, IT 22, which is derived from 3T3 cells²⁴. The reaction with this protein suggests that some anti-*abl* sera recognize antigenic determinants unique to p250 but that others do not, as it is only detected with certain anti-*abl* sera (J.B.K. and O.N.W., unpublished results). No p210 was precipitated by normal rabbit serum (Fig. 3a).

We have used somatic cell hybrids between LM-TK⁻ mouse fibroblasts and cell lines K562 and Bv173 derived from CML patients to study the effect of the t(9; 22) translocation on the

Table 1 Characterization of hybrid clones

Hybrid clones	Oncogene present		C_A	Relevant chromosomes*
	<i>c-abl</i>	<i>c-sis</i>		
KL-17-9-1	+	+	+	Ph ¹ , 22
KL-17-9-3	+	+	+	Ph ¹ , 22
260-3-1-3	+	—	+	Ph ¹

The DNA of hybrid clones was screened by Southern blot analysis using the *v-abl*, C_A and *c-sis* probes. Cells that were positive for *c-abl* and C_A and negative for *c-sis* contained the Ph¹ chromosome. Hybridization with *c-sis* indicated the presence of either normal chromosomes 22 and 9, or chromosomes 9q+ and Ph¹. Isoenzyme analysis for AK-1 and AK-3, markers for the long and short arms of chromosome 9, respectively revealed no normal chromosome 9 or 9q+ chromosome present in the hybrids.

* Data are based on Southern blot and isoenzyme analyses.

regulation of *c-abl* expression. As the altered *c-abl* RNA and protein are present in hybrid cells which contain only the Ph¹ chromosome and not other relevant chromosomes, we conclude that they must be derived from the *c-abl* locus on the Ph¹ chromosome. The results of the analysis of *c-abl* protein expression in these hybrids are consistent with the proposed relationship between the transcript and the p210 *c-abl* protein¹¹.

We have reported previously that there is a marked reduction in the levels of human *c-myc* transcripts in hybrid cells of mouse fibroblasts and Burkitt's lymphoma cells possessing the 8;14 translocation²⁵⁻²⁷. Thus, although the translocated *c-myc* oncogene is expressed at elevated levels in the B-cell background, it is clearly repressed in cells of other differentiation lineages (for example, fibroblasts)²⁵⁻²⁷. In contrast, the expression of the *c-abl* 8-kb transcripts and p210 in the fibroblast hybrids indicates that the activation of the *c-abl* oncogene as a result of the chromosomal translocation is not myeloid-specific. Furthermore, the presence of the Ph¹ chromosome in haematopoietic lineages other than myeloid^{28,29} raises the possibility that the effect of the altered *c-abl* protein differs for different cell types. This possibility is also supported by the observation that Epstein-Barr virus-transformed lymphoblastoid cell lines derived from Ph¹-positive CML patients express the 8-kb aberrant *c-abl* transcripts and the p210 protein (O.N.W., unpublished results).

We thank Ms Marina Hoffman for editorial assistance. This work was supported by NCI grants CA-16685, CA-36521, CA-23568, CA-25875 to C.M.C., NIH grant GM-32583 to L.C.S., and by a Faculty Scholar Award from the American Cancer Society and by NIH research grants CA-27505, CA-32737 to O.N.W. D.K. is a postdoctoral fellow of the National Cancer Institute of Canada. J.B.K. is supported by USPHS National Service Award GM-07185.

Received 25 March; accepted 23 October 1985.

- Nowell, P. C. & Hungerford, D. A. *J. natn. Cancer Inst.* **25**, 85-109 (1960).
- Rowley, J. D. *Nature* **243**, 290-291 (1979).
- Rowley, J. D. *Science* **216**, 749-751 (1982).
- de Klein, A. *et al. Nature* **300**, 765-767 (1982).
- Dalla-Favera, R., Gallo, R. C., Giallongo, A. & Croce, C. M. *Science* **218**, 686-688 (1982).
- Groffen, J. *et al. J. exp. Med.* **158**, 9-15 (1983).
- Cannani, E. *et al. Lancet* **i**, 593-595 (1984).
- Collins, S. J., Kubonishi, I., Miyoshi, I. & Groudine, M. T. *Science* **225**, 72-74 (1984).
- Gale, R. P. & Canaani, E. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5648-5652 (1984).
- Konopka, J. B., Watanabe, S. M. & Witte, O. N. *Cell* **37**, 1035-1042 (1984).
- Davis, R. L., Konopka, J. B. & Witte, O. N. *Molec. cell. Biol.* **5**, 204 (1985).
- Konopka, J. B., Watanabe, S. M., Singer, J. W., Collins, S. J. & Witte, O. N. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1810-1814 (1985).
- Dubbs, D. R. & Kit, S. *Exp. Cell Res.* **33**, 19-28 (1964).
- Lozzio, C. B. & Lozzio, B. B. *Blood* **45**, 321-334 (1975).
- Peglararo, L. *et al. J. natn. Cancer Inst.* **70**, 447-453 (1983).
- Srinivasan, A., Reddy, E. P. & Aaronson, S. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2077-2081 (1981).
- Croce, C. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 6922-6926 (1983).
- Heiter, P. A., Hollis, G. F., Korsmeyer, S. J., Waldman, T. A. & Leder, P. *Nature* **296**, 536-560 (1981).
- Collins, S. J. & Groudine, M. T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4813-4817 (1983).
- Selden, J. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 7289-7292 (1983).
- Tsujimoto, Y. & Croce, C. M. *Nucleic Acids Res.* **12**(22), 8407-8414 (1984).
- McKusick, V. A. *Cytogenet. Cell Genet.* **32**, 7 (1982).

- Ruddle, R. & Creagan, R. A. *Rev. Genet.* **9**, 407-486 (1975).
- Croce, C. M., Barrick, J., Linnenbach & Koprowski, H. *J. cell. Physiol.* **99**, 279-284 (1979).
- Nishikura, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 4822-4826 (1983).
- Nishikura, K. *et al. Science* **224**, 399-402 (1984).
- Croce, C. M., Erikson, J., ar-Rushdi, A., Aden, D. & Nishikura, K. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3170-3174 (1984).
- Karpas, A., Fischer, P. & Swirsky, D. *Lancet* **i**, 931-933 (1982).
- Minowada, J., Tsubota, T., Greaves, M. F. & Walters, T. R. *J. natn. Cancer Inst.* **70**, 447-453 (1983).

Disease-specific and tissue-specific production of unintegrated feline leukaemia virus variant DNA in feline AIDS

James I. Mullins*, Carol S. Chen* & Edward A. Hoover†

* Department of Cancer Biology, Harvard School of Public Health, Boston, Massachusetts 02115, USA

† Department of Pathology, College of Veterinary Medicine, Colorado State University, Ft Collins, Colorado 80523, USA

Feline leukaemia viruses (FeLVs) have long been known to be associated with induction of proliferative and anti-proliferative diseases of domestic cats¹. Strains of FeLV have been recognized which specifically induce lymphosarcoma^{2,3}, aplastic anaemia⁴⁻⁶, myelodysplastic anaemia⁶, and, recently, feline AIDS (acquired immune deficiency syndrome)⁷ strikingly similar to human AIDS which is lethal in 100% of inoculated and viraemic specific-pathogen-free (SPF) cats⁸. Here, we have analysed FeLV DNA in tissues of 22 SPF cats that had been inoculated with the feline AIDS strain (FeLV-FAIDS) and we find two classes of viral DNA—a monotypic common form which is detectable in bone marrow regardless of disease state, and variant forms, recognizable by restriction site differences, whose appearance correlates with onset of disease symptoms and persists throughout the course of the disease. FeLV-FAIDS variant DNA is detected at high concentration (10–50 copies per cell) and principally as unintegrated viral DNA (UVD) in bone marrow of cats with feline AIDS. In marked contrast, high levels of UVD were not present in cats in the terminal-stages of T-cell lymphosarcoma, aplastic anaemia, or myelodysplastic anaemia induced by other FeLV strains. These results parallel recent observations in humans, where high levels of UVD were sometimes found in cells derived from AIDS patients infected with human T-lymphotropic virus type III (HTLV-III)/lymphadenopathy-associated virus (LAV), and suggest that persistence of unintegrated variant viral DNA is a crucial indicator of retrovirus-induced cytopathic disease syndromes such as AIDS.

DNA from the tissue samples described below was extracted and analysed for the presence of exogenous FeLV sequences after digestion with the restriction endonuclease *KpnI*, agarose gel electrophoresis, Southern transfer and hybridization using the exogenous FeLV long terminal repeat (LTR)-specific hybridization probe exU3 as described previously⁹. In this analysis, viral DNA produces two bands which hybridize to the exU3 probe: an internal virus band derived from the 3' end of the viral genome and a band derived from LTR sequences at the 5' end of the genome, which, if linked to adjacent cellular DNA, is referred to as a host-virus junction band. The intensity of the internal band reflects the average provirus copy number per cell in the tissue examined. The DNA in Fig. 1, lane 1, was derived from RD-FeLV-2, a single cell clone of human RD cells infected with the Gardner-Arnstein subgroup B strain of FeLV (FeLV-B-GA)¹⁰. A 3.65-kilobase (kb) internal virus band with an intensity representing ~22 copies was detected, as well as an equivalent number of single-copy bands derived from host-virus junction fragments. DNA derived from a field case of T-cell lymphosarcoma (animal DQ; Fig. 1, lane 2) gave a similar

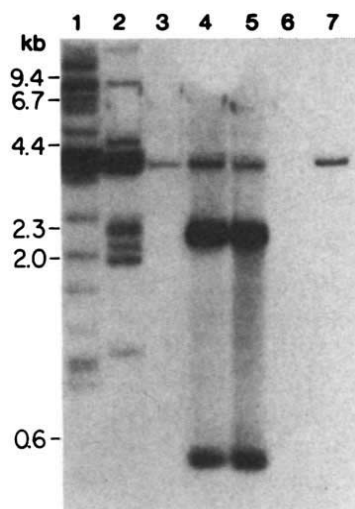


Fig. 1 Exogenous FeLV sequences in animals infected with FeLV-FAIDS. The DNA samples analysed were the RD-FeLV-2¹⁰ cloned cell line (lane 1), thymic tumour from animal DQ (lane 2), spleen (lane 3) and bone marrow (lane 4) from animal 1082 with experimentally induced feline AIDS, bone marrow (lane 5) and lymph node (lane 6) from animal 1220 with experimentally induced feline AIDS, and bone marrow (lane 7) from animal 1044, an asymptomatic animal inoculated with FeLV-FAIDS. Relative molecular mass (M_r) markers are shown in kilobase pairs (kb).

Methods. Total cellular DNA (7.5 μ g) was purified, digested with *Kpn*I, electrophoresed in 1.2% agarose gel, transferred to nitrocellulose and hybridized in the presence of 10% dextran sulphate and 50% formamide as described previously⁹. The probe, derived by gel purification²³ of the insert in plasmid pexU3 (J.I.M. *et al.*, in preparation), was ³²P-nick-translated²⁴ to a specific activity of 8×10^8 d.p.m. μ g⁻¹ and was used at a concentration of 10^6 c.p.m. ml⁻¹. After hybridization at 42 °C, for 36 h, the filter was incubated in five changes of $2 \times$ SSC (0.3 M NaCl, 0.03 M sodium citrate), 0.05% sodium pyrophosphate and 0.1% sodium lauryl sarkosine at 50 °C, dried and exposed to X-ray film at -70 °C with an intensifying screen for 6 h prior to development. Mean virus copy numbers (per cell) were estimated from densitometer scans of the exposed films.

pattern, with a single prominent 3.65-kb internal virus band and ~12 single-copy host-virus junction bands, indicating that the tumour is a clonal cell outgrowth of FeLV-infected cells.

In the present study, cell-free tumour extract from DQ was used as a source of the feline AIDS virus⁸ and was inoculated into newborn, weanling and adult cats. Two cats (1220, 1082) that developed acute feline AIDS following inoculation were killed and their tissue DNA prepared for analysis. Spleen DNA from cat 1082 showed a non-clonal pattern of virus integration, with a single copy of the internal virus band of 3.65 kb, the same length as the prominent provirus in the DQ tumour (Fig. 1, lane 3). The bone marrow of both diseased cats contained ~2 copies per cell of this provirus (Fig. 1, lanes 4 and 5) but in addition contained two bands of intensity ~15 copies per cell, one at 2.1 kb and the other at 0.4 kb. No viral DNA was detected in the lymph node of cat 1220 (Fig. 1, lane 6), who had severe lymphoid cell depletion in that organ. Bone marrow from an adult cat (1044) that was asymptomatic when the sample was taken, contained only the form of virus giving the 3.65-kb *Kpn*I band (Fig. 1, lane 7).

The state of viral DNA integration was examined by Southern blot analysis without prior restriction enzyme digestion (Fig. 2). A prominent band at 8.7 kb was detected in each feline AIDS bone marrow sample (Fig. 2, lanes 3, 5), co-migrating with unintegrated lines FeLV DNA derived from RD cells freshly infected with FeLV-B-GA (Fig. 2, lane 4)¹⁰. A deleted 7.5-kb form of unintegrated viral DNA and a small amount of relaxed circular viral DNA migrating with 18-kb linear DNA^{10,11} were

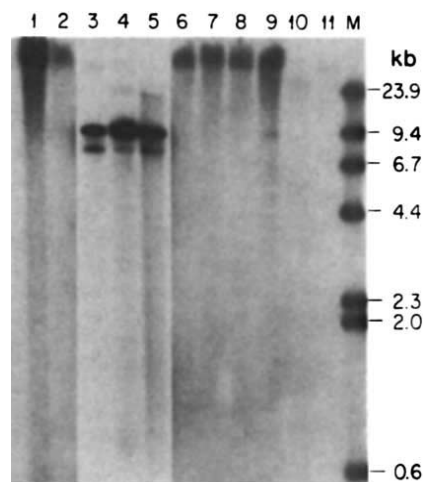


Fig. 2 Exogenous FeLV in undigested cat cellular DNAs. Total cellular DNA was purified and analysed as described in Fig. 1 legend, except that samples were examined without restriction enzyme digestion and a 0.8% agarose gel was used. DNA samples analysed were: lane 1, thymic tumour from animal DQ; lane 2, spleen from animal 1082; lane 4, the extrachromosomal, Hirt-supernatant fraction²⁵ of RD cell DNA taken 24 h after infection with FeLV-B-GA at a multiplicity of infection ~1.0 ref. 10); lanes 3 and 5, bone marrow samples from animals 1082 and 1220, respectively, lanes 6, 7, two animals with aplastic anaemia; lane 8, animal 1044; lane 9, an animal with a thymic tumour; lanes 10, 11, two animals with myelodysplastic anaemia. Lanes 3-6 were exposed to X-ray film for 6 h before development; the remaining lanes were exposed for 16 h. Lane M, M_r markers.

detected in both the RD infected (FeLV-B-GA) cell preparation and the feline AIDS bone marrow samples. Evidently, over 90% of viral sequences in these cells are present as UVD and it is unclear whether any bone marrow-specific viral sequences are integrated into cell DNA. As it is unlikely that all cells within the bone marrow are infected and/or have the same amount of unintegrated DNA, we suspect that the concentration of UVD in some infected cells is much higher than the average of 15 variant molecules per cell noted above.

A similar analysis was performed 'blindly' on bone marrow samples taken from 19 additional animals inoculated with FeLV-FAIDS and a sample from animal 1044 following the development of symptoms of feline AIDS. Representative examples of this survey are shown in Fig. 3. In the bone marrow from animal 1044 and from an additional animal (1161) with overt disease, we detected ~50 copies per cell of unintegrated FeLV-variant DNA. We found a low level of integrated variant DNA in animal 1276, which developed clinical signs of acute disease within a week after the bone marrow sample was extracted for analysis. Terminal tissue samples from this animal revealed a high level of variant UVD in bone marrow and common as well as variant proviral DNA in other tissues (data not shown).

Unintegrated viral DNA is an obligatory intermediate in the life cycle of retroviruses¹², it is detected *in vitro* usually only for a few days following synchronous infection of cells at high virus multiplicity, as shown in Fig. 2 for FeLV-B-GA-infected RD cells. UVD is rarely detected *in vivo* as few cells in a tissue are likely to have been infected immediately before tissue sampling. Known exceptions to this include a recent report of persistent UVD in the bone of avian leukosis virus-infected osteopetrotic chickens¹³ and following generation of MCF recombinant viruses in pre-leukaemic AKR mice¹⁴. Interestingly, transient production of a high level of UVD following *in vitro* infection with certain avian retroviruses correlates with viral cytopathicity^{15,16}, and infection of sheep choroid plexus cells with the cytopathic ovine visna lentivirus results in persistent synthesis of UVD¹⁷. Therefore, we sought to determine whether

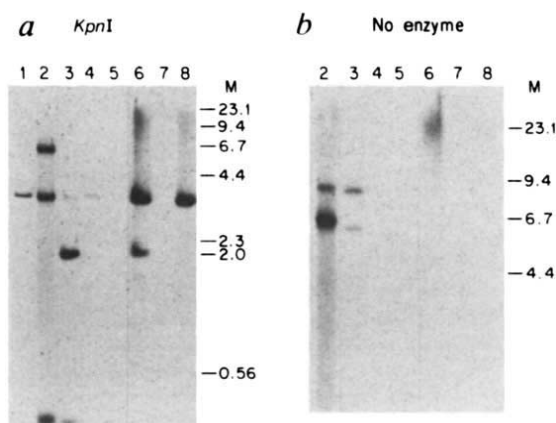


Fig. 3 Exogenous FeLV in animals inoculated with FeLV-FAIDS. DNA samples derived from bone marrow of 19 animals were analysed as described in Fig. 1 legend (a, *KpnI* digests) and as described in Fig. 2 legend (b, undigested DNA). Samples from seven animals and RD-FELd-2 cells (lane 1) are shown: those in lanes 2 and 3 are from animals (1044 and 1161, respectively) with feline AIDS. Animal 1276 (lane 6) developed symptoms of feline AIDS within 1 week after the bone marrow sample was taken. Samples in lanes 4 (1237), 5 (1238) and 8 (1283) are from viraemic, asymptomatic animals. The sample in lane 7 (1282) is from an asymptomatic animal in which viral infection regressed.

detection of UVD in cats was specific to feline AIDS and not found in the bone marrow of cats with other diseases induced by defined strains of FeLV (Fig. 2). DNAs from the following sources were analysed for the presence of UVD: terminal-stage thymic lymphosarcoma tissue from cat DQ (Fig. 2, lane 1); spleen from cat 1082 with feline AIDS (lane 2); bone marrow from two SPF cats with terminal-stage aplastic anaemia induced experimentally by the molecularly cloned subgroup C, Sarma strain of FeLV (FeLV-C-S)⁶ (lanes 6, 7); bone marrow from a viraemic, asymptomatic cat inoculated with FeLV-FAIDS (cat 1044, lane 8; 17 more cats in this category were examined with identical results); bone marrow from a tumour-bearing cat inoculated with the T-cell lymphosarcoma-inducing Rickard strain of FeLV² (lane 9); and bone marrow from two SPF cats with terminal-stage myelodysplastic anaemia induced by inoculation of molecularly cloned B-FeLV-GA (J.I.M. *et al.*, in preparation) (lanes 10, 11). We detected no UVD in the DQ tumour or bone marrow of any other animal, except at a low concentration in the marrow from the cat with lymphosarcoma induced by FeLV-Rickard (less than one copy per cell in this clonal cell population).

Because of the striking similarity between the pathology of human and feline AIDS, we sought to determine whether HTLV-III/LAV, the aetiological agent of human AIDS¹⁸, could be detected as unintegrated viral DNA. We (N. Riedel *et al.*, in preparation) and others¹⁹ found that 25% of the HTLV-III/LAV DNA present in the established cell line HT/H9 was present as unintegrated linear viral DNA. Shaw *et al.*¹⁸ also reported detection of UVD in a few fresh tissues from AIDS patients, whereas we detected no UVD in any of four HTLV-III/LAV DNA-positive fresh tissue samples.

We detect three features in the behaviour of FeLV-FAIDS which distinguish it from all previously recognized FeLV strains: (1) marked tissue-specific replication of a variant form of virus; (2) distinct disease-specific association with synthesis of variant FeLV DNA; and (3) high concentrations of unintegrated viral DNA *in vivo* at terminal stages of the disease. The possibility that feline AIDS variant FeLVs possess a cell-specific tropism is supported by our finding of high concentrations of variant viral DNA in bone marrow but not at detectable levels in other tissues in two out of three cats examined. It is unclear at present whether cell- or tissue-selective replication, *de novo* recombination or mutation accounts for the detection of the bone marrow-

specific feline AIDS variant. As every cat that became viraemic after inoculation with the DQ tissue-derived stock of FeLV-FAIDS developed severe immunosuppression and clinical feline AIDS⁸ and four of five variants detected had the same 2.1 kb internal *KpnI* fragment, it is possible that variant viruses were present but did not replicate efficiently in the initial lymphosarcoma tissue. The correlation between the production of variant UVD and the appearance of disease is particularly striking. Each of five cats with overt disease had high levels of variant UVD in the bone marrow, whereas none of 17 FeLV-FAIDS viraemic but asymptomatic cats had detectable levels of variant or UVD in their bone marrow.

The FeLV-FAIDS preparation contains viruses which fall within the A and B interference subgroups (ref. 20 and O. Jarrett, personal communication). Subgroup A viruses are generally considered to be minimally pathogenic but able to establish viraemia in a high proportion of infected cats^{20,21}. Subgroup B viruses are less capable of establishing viraemia in other than newborn animals without concurrent infection with subgroup A FeLV²¹. The appearance of FeLV-B in the blood of cats infected with both subgroups has been associated with onset of other FeLV-associated diseases²¹, and at least one strain of molecularly cloned FeLV-B is pathogenic when inoculated into newborn SPF cats (J.I.M. *et al.*, in preparation). Since the bone marrow-specific form of FeLV-FAIDS was detected as unintegrated linear viral DNA and because persistence of high levels of UVD is associated with retrovirus cytopathicity, we suggest that the bone marrow-specific component is likely to be the cytopathic component of the strain mixture (possibly the subgroup B component, by analogy with the work of Jarrett and co-workers²¹) while the subgroup A component is probably the highly replicative, helper-like, common form virus.

The striking similarity between the clinico-pathological features of human and feline AIDS⁸, the apparent cytopathic effect of both viruses on selected T-cell populations⁸, and the detection of persistent unintegrated viral DNA in both disease processes suggest that experimentally induced feline AIDS will serve as a useful model in which to: (1) investigate the pathogenetic mechanisms of AIDS; (2) delineate the viral genomic component(s) responsible for lymphocytopenia; and (3) evaluate the efficacy of antiviral compounds or viral vaccines in the treatment or prevention of retrovirus-mediated acquired immune deficiency in humans as well as animals. Unlike LAV/HTLV-III, which is not closely related to HTLV-I and -II²², the feline AIDS virus is closely related to other FeLVs (for example, detection of virus using the exU3 hybridization probe indicates strong conservation of LTR-enhancer sequences). The relatedness of apathogenic FeLVs and FeLV-FAIDS strains should facilitate elucidation of the viral determinants critical to immunosuppressive cytopathicity.

We thank Dr M. O. Nicolson for providing Hirt-supernatant FeLV DNA. This work was supported by grants to J.I.M. from the NIH (CA34975) and the Massachusetts AIDS Research Council, and by a scholar award from the American Cancer Society, Massachusetts Division, and by grants to E.A.H. from the NIH (CA32552 and CA32563).

Received 21 June; accepted 1 November 1985.

- Hardy, W. D. Jr. in *Feline Leukemia Virus* (eds Hardy, W. D. Jr, McClelland, A. J. & Essex, M.) 3-78 (Elsevier, New York, 1980).
- Rickard, C. G., Post, J. E., de Noronha, F. & Barry, L. M. *J. natn. Cancer Inst.* **42**, 987-1014 (1969).
- Hoover, E. A., Perryman, L. E. & Kociba, G. J. *Cancer Res.* **33**, 145-152 (1973).
- Hoover, E. A. *et al.* *J. natn. Cancer Inst.* **53**, 1271-1276 (1974).
- Mackey, L., Jarrett, W., Jarrett, O. & Laird, H. *J. natn. Cancer Inst.* **54**, 209-217 (1975).
- Riedel, N., Hoover, E. A., Gasper, P. W., Nicolson, M. O. & Mullins, J. I. *J. Virol.* (submitted).
- Hardy, W. D. in *Human T-Cell Leukemia/Lymphoma Virus* (eds Gallo, R. C., Essex, M. E. & Gross, L.) 35-43 (Cold Spring Harbor Laboratory, New York, 1984).
- Hoover, E. A., Mullins, J. I., Gasper, P. W. & Quackenbush, S. L. *Science* (submitted).
- Mullins, J. I., Brody, D. S., Binari, R. C. Jr & Cotter, S. M. *Nature* **308**, 856-858 (1984).
- Mullins, J. I. *et al.* *Nucleic Acids Res.* **8**, 3287-3305 (1980).
- Sherr, C. J. *et al.* in *Feline Leukemia Virus* (eds Hardy, W. D. Jr, McClelland, A. J. & Essex, M.) 293-307 (Elsevier, New York, 1980).
- Varmus, H. & Swanstrom, R. in *RNA Tumour Viruses* (eds Weiss, R., Teich, N., Varmus, H. & Coffin, J.) 369-512 (Cold Spring Harbor Laboratory, New York, 1982).

13. Robinson, H. L. & Miles, B. D. *Virology* **141**, 130-143 (1985).
14. Herr, W. & Gilbert, W. *J. Virol.* **50**, 155-162 (1984).
15. Keshet, E. & Temin, H. M. *J. Virol.* **31**, 376-388.
16. Weller, S. K. & Temin, H. M. *J. Virol.* **39**, 713-721 (1981).
17. Harris, J. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 7212-7215 (1984).
18. Gallo, R. C. *et al. Science* **224**, 500-503 (1984).
19. Shaw, G. M. *et al. Science* **226**, 1165-1171 (1984).
20. Sarma, P. S. *et al. Proc. Soc. exp. Biol. Med.* **14**, 757-762 (1974).
21. Jarrett, O. in *Feline Leukemia Virus* (eds Hardy, W. D. Jr, McClelland, A. J. & Essex, M.) 473-479 (Elsevier, New York, 1980).
22. Gonda, M. A. *et al. Science* **227**, 173-177 (1985).
23. Waalwijk, C. & Flavell, R. A. *Nucleic Acids. Res.* **5**, 3231-3236 (1978).
24. Mullins, J. I. *et al. J. Virol.* **38**, 688-703 (1981).
25. Hirt, B. *J. molec. Biol.* **26**, 365-369 (1967).

Structural homology of the product of the *Drosophila Krüppel* gene with *Xenopus* transcription factor IIIA

Urs B. Rosenberg, Christian Schröder, Anette Preiss, Andrea Kienlin, Serge Côté, Isolde Riede* & Herbert Jäckle

Max-Planck-Institut für Entwicklungsbiologie, Abteilung Biochemie II, Spemannstrasse 35, D-7400 Tübingen, FRG

* Max-Planck-Institut für Biologie, Abteilung Henning, Corrensstrasse 38, D-7400 Tübingen, FRG

Segmentation of the *Drosophila* embryo is established at around the blastoderm stage¹⁻³ and requires both maternal information in the egg cytoplasm^{4,5} and expression of the zygotic genome. Zygotic genes involved in segmentation have been defined by mutations that affect the segmental pattern of the embryo⁶⁻¹⁰, and these fall into several different classes, such as the 'gap' genes, mutations of which cause the loss of contiguous segments⁶. *Krüppel* (*Kr*) is an example of a *Drosophila* gap gene, strong *Kr* mutant embryos lacking all thoracic and five anterior abdominal segments, with part of the remaining posterior abdominal segments being present as a mirror-image duplication (weaker alleles cause shorter deletions^{11,12}). *Kr* has been cloned¹² and shown to encode a blastoderm-gastrulation stage-specific transcript expressed in regions of the embryo affected by *Kr* mutants¹³. We report here that the protein product of the *Kr* gene predicted from DNA sequences is structurally homologous to the transcription factor TFIIIA which regulates 5S gene expression in *Xenopus*. Like TFIIIA, the predicted *Kr* protein has an internally repetitious sequence, and as has been suggested for TFIIIA, the repeat units may form DNA-binding domains.

Different DNA fragments from the 18-kilobase (kb) DNA segment containing essential *Kr*⁺ sequences (-5.1 to +12.7; for details see Fig. 3 in ref. 12) were used to analyse Northern blots loaded with RNA from different stages of *Drosophila* development (Fig. 1). None of the fragments hybridized to poly(A)⁻ RNA (data not shown) and only one, the 5.4-kb fragment (+1.5 to +6.9), hybridized to poly(A)⁺ RNA in a broad 2.5-kb band, as seen earlier with an overlapping 9.5-kb *Eco*RI fragment¹². A series of corresponding complementary DNA clones were isolated from several libraries prepared from poly(A)⁺ RNA of 2-5-h-old embryos (blastoderm to gastrula). All of them map within a 2.5-kb segment of genomic DNA (Fig. 1a,b).

Different fragments from two almost full-size cDNA clones (see Fig. 1b) were probed to developmental Northern blots. Each of the fragments hybridized to the broad 2.5-kb band of poly(A)⁺ RNA seen with the genomic DNA fragment (see above). Using higher-resolution agarose gels (Fig. 1c,h), this band could be resolved as a doublet of transcripts lying about 350 base pairs (bp) apart. Both transcripts were expressed in 2-5-h-old embryos (blastoderm-gastrula). Extended exposure of the Northern blots revealed that *Kr*⁺ transcripts are present up until larval stages, being about two orders of magnitude below their abundance at blastoderm-gastrulation stages (data not shown).

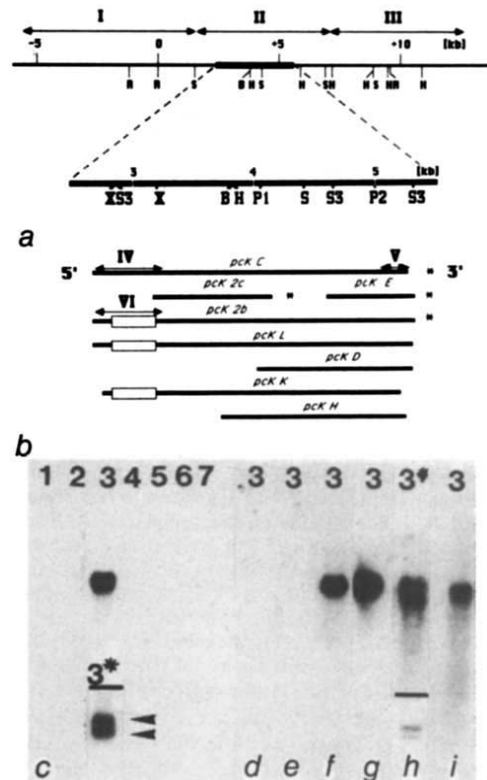


Fig. 1 *Kr* transcript analysis. *a*, Restriction map of the *Kr* region (for details see Fig. 3 in ref. 12) and its enlarged core (below). Restriction enzymes used are *Eco*RI (R), *Sal*I (S), *Bam*HI (B), *Hind*III (H), *Pvu*I and *Pvu*II (P1, P2), *Xba*I (X) and *Sau*3A (S3). Fragments I, II and III, which originate from the ER3 clone¹², were used as probes for Northern analysis (lanes c-e). *b*, Alignment of *Kr* cDNA clones relative to the restriction map of genomic DNA shown above. Boxes indicate an intron (see Fig. 2) that is absent from the respective clones. The 5' to 3' orientation was established by single-strand cDNA mapping³¹ and confirmed by sequencing (see Fig. 2) of genomic DNA and the cDNA clones marked with an asterisk. Note the location of fragments IV, V and VI used for Northern analysis. Lanes c-i, localization of *Kr* transcripts within cloned DNA and the developmental profile of *Kr* transcript accumulation. Numbers over the lanes designate the developmental stages at which the poly(A)⁺ RNA was extracted: 1, unstaged ovaries; 2, 0-1-h-old embryos; 3 and 3*, 2-5-h-old embryos; 4, 8-24-h-old embryos; 5, larvae; 6, pupae; 7, adults. *c*, Developmental Northern profile of *Kr* transcripts. Note that the broad 2.5-kb poly(A)⁺ RNA band (lanes 3) is actually a doublet of bands (lane 3*) which differ in ~350 bp. The Northern blot was probed with fragment II. The same Northern blot (only lanes 3 shown) was probed with fragments I (lane d) or fragment III (lane e). In both cases, extended exposure revealed no signal. Parallel Northern blots were probed with pcK2b (lane f) and pcKC (lane g). Lane h, Northern blots probed with intron containing (probe IV, upper portion) and intron lacking (probe VI, lower portion) sequences prepared from cDNA clones. Note the similar intensities of the doublets with probe VI and a stronger signal of the longer band with probe IV. Lane i, Northern blot probed with the 3' end of one cDNA clone (probe V). Note the low level of background smear due to poly(A) tail homologies.

Methods. cDNA clones were isolated from libraries prepared from 2-4-h and 2-5-h embryos. pcK2b and pcK2c were found in the Goldschmidt-Clermonts library (unpublished); the other clones were from a freshly prepared library (U.B.R., unpublished). Preparation of poly(A)⁺ RNA, nick-translation of the probes and Northern blot analysis have been described earlier¹². Gels for Northern blots were loaded with 10 µg poly(A)⁺ RNA per lane using 1.0% agarose gels (lanes without asterisks) or 2 µg poly(A)⁺ RNA using 0.7% agarose gels and prolonged separation (lanes with asterisks).

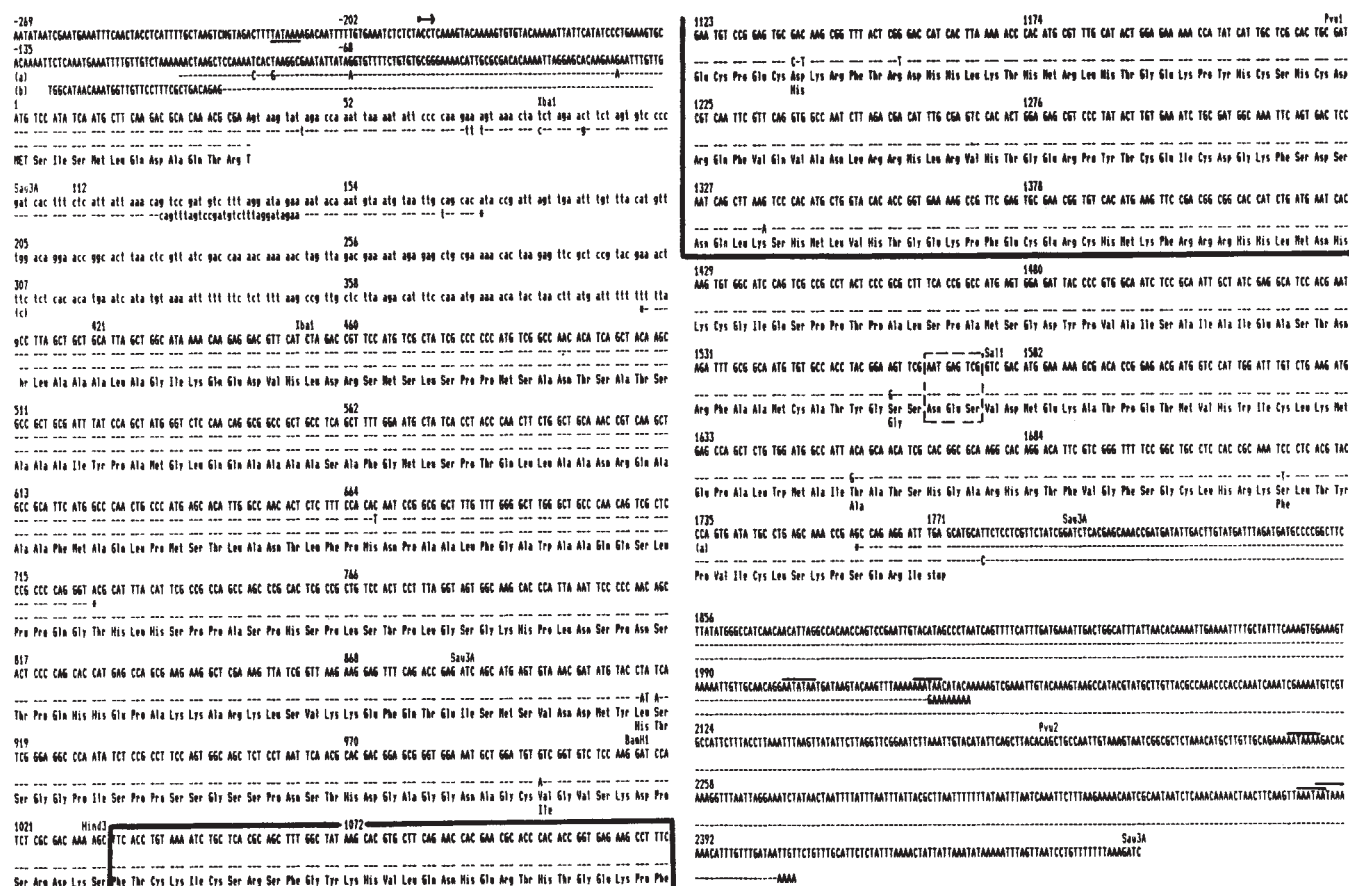


Fig. 2 Sequence of the *Kr* gene and presumptive amino-acid sequence of the *Kr* protein. Sequence of *Kr* genomic DNA, cDNA sequences (pcKC, pcK2b and pcK2c designated (a), (b) and (c) below the genomic sequence and the only open reading frame amino-acid sequence below. Asterisks indicate that cDNA clone pcKC is not sequenced completely; the intron acceptor site was sequenced in pcKC (sequence identical to genomic DNA) and in pcK2c (between the second and third asterisks in the cDNA pcKC sequence). Note the TATA box (underlined), the cap site (arrow), the exon/intron/exon borders in positions 37 and 409, a presumptive glycosylation site at position 1,564–1,572, the translation stop in position 1,771–1,774 followed by a series of five polyadenylation signals (overlined). Note that at least two of them are used in pcK2b and pcKC. The heavily boxed sequence (position 1,036–1,371) contains the domain of the four complete 28-amino-acid repeat sequences followed by a fifth incomplete repeat unit (see text and Fig. 3 for details). For sequencing, genomic DNA clones and the cDNA clones (see Fig. 1b, labelled by asterisks) were subcloned into pUC 18/19 or mp 18/19 using a single restriction enzyme (*Xba*I or *Sau*3A) or a combination of the others (see Fig. 1a) to produce subclone fragments in overlapping orientations. Dideoxy sequencing was carried out as described previously¹⁵, using both sequencing of both strands and verification of the sequence in cDNA clones as a strategy. In this way, each part was confirmed in a minimum of two independent sequencing reactions.

The *Kr* gene contains a 1,398-bp open reading frame which is split by a 372-bp intron (Fig. 2). Transcription starts at position –185, consistent with both primer extension studies (data not shown) and the structure of the cDNA clones (see below). Two 'TATA' boxes occur 86 and 26 bp upstream from this transcriptional start; the latter (underlined in Fig. 2) is the more conventional spacing. A series of five polyadenylation signals are found downstream from the translated sequence. The first Met codon in our sequence starts with residue 184 downstream from the transcription start. However, neither this nor a second Met codon upstream from the intron border fits the consensus sequence (CCG/ACCATGG) for eukaryotic translation initiation sites¹⁴ and, therefore, initiation of translation at the first conventional Met codon (in position 466–469) cannot be positively ruled out.

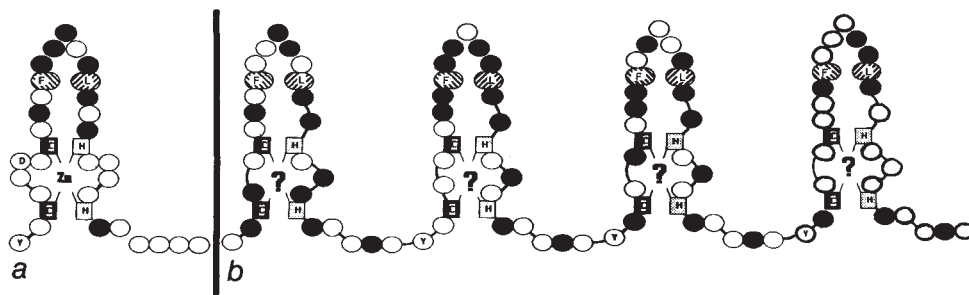
One *Kr* cDNA clone (pcK2b) was sequenced completely, others were partly sequenced (Fig. 2). The cDNA clones pcK2b and pcKC terminate at different positions, suggesting that more than one of the five potential polyadenylation sites are used. The 5' end of the pcKC cDNA sequence starts at position –101, indicating that this clone is close to full size. In contrast, the 5' end of the pcK2b cDNA clone does not correspond to the genomic sequence, suggesting an artefact during cDNA synthesis (as observed earlier in other cDNAs of the same library; for details see refs 15, 16). The major difference between the

two cDNA clones, aside from using a different polyadenylation site (see above), is that the intron is spliced in pcK2b but present in pcKC and several other cDNA clones (Fig. 1b). Nucleotide differences, including a polymorphic *Xba*I site (position 82–87), a small duplicated segment in the intron and a few base pair changes, reflect minor differences in various wild-type strains (data not shown) and mismatch reading by the reverse transcriptase during cDNA construction.

Translation initiation at the first ATG in the *Kr* transcript (position 1) results in a small 19-amino-acid peptide which is interrupted by a stop codon inside the intron present in clone pcK2b. However, initiation at the same position results in a readthrough equivalent to a 466-amino-acid protein (Fig. 2) in spliced *Kr* messenger RNA represented by clone pcK2b. Alternatively, translation could be initiated at the 'conventional' ATG codon in position 484 (ref. 14), leaving an unusually long 5'-untranslated sequence in both transcripts for possible regulation at the translational level.

The *Kr* protein deduced from the 1,398-bp open reading frame seems to be divided into three domains. The amino-terminal region up to position 1,035 is extremely rich in Ala, Ser and Pro (>30%) and folds into a globular structure. The carboxy-terminal region (after position 1,480) shows similar features, including a putative glycosylation site¹⁷. In contrast to

Fig. 3 Folding scheme for a linear arrangement of a TFIIIA repeated domain and *Kr* protein. **a**, TFIIIA 'finger structure' as described in detail in ref. 18. Each of the nine repeated domains is centred on a tetrahedral arrangement of zinc (Zn) ligands. Marked amino-acid residues are conserved. They include Tyr (Y), the Cys (C) and His (H) zinc ligands and the hydrophobic Phe (F) and Leu (L). Black circles mark the position of DNA-binding amino acids²¹.



The folding scheme shows that zinc draws the ends of each unit together, leaving the central residues to form a potential DNA-binding loop (finger). **b**, *Kr* protein folding structure. Note the structural homology to the proposed TFIIIA structure, leaving the Cys-His arrangement and central residues to form a potential DNA-binding finger. Note that His 23 is followed by an invariant stretch of amino acids (position 24–28), and the variable positions of DNA-binding amino acids (black circles) in the finger loop (position 8–19). For the amino-acid sequence see the boxed area in Fig. 2. Orientation of the protein domain is amino (left) to carboxy terminus (right). Since the fifth, incomplete repeat unit lacks His 23, it may not fold into a finger and is therefore not included in the scheme.

the central portion (see below), the amino and carboxyl ends are relatively neutral (10% basic). At position 1,036, the *Kr* protein starts an internal 28-amino-acid repeat unit consisting of the following motif: Phe 1 (or Tyr 1), Cys 3, Cys 6, Phe 10, Leu 16, His 19, His 23 followed by Thr, Gly, Glu, Lys or Arg (position 24–27) and Pro 28 (see Figs 2, 3). Four such complete units are arranged tandemly, ending with Pro 28 in position 1,372, and they are followed by a fifth incomplete repeat which shares most of the above structural motif. This portion of the *Kr* protein domain may fold into a structure similar to that proposed by Klug and co-workers¹⁸ for TFIIIA (see Fig. 3); TFIIIA is needed for transcription by polymerase III, which acts on a limited number of genes including those coding for 5S RNA of *Xenopus laevis*. The TFIIIA protein was shown to bind both 5S RNA and its cognate DNA and represents a key element in the autoregulation of 5S gene transcription^{19,20}. In analogy to TFIIIA, the *Kr* repeat unit may fold into four potential DNA-binding loops ('fingers'), each centred on invariant pairs of Cys and His (compare Fig. 3a and b). Each finger consists of two hydrophobic amino acids (Phe 10, Leu 16) surrounded by non-conserved, mostly basic or positively charged amino acids, followed by a stretch of five conserved amino acids (position 24–28). The 'fingertip' is rich in positively charged 'DNA-binding'²¹ amino acids, representing the potential DNA-binding region of the *Kr* protein. The specificity of DNA binding would be provided by the side groups of Lys, His, Asn, Gln and Thr (interacting with the phosphate backbone of DNA) and Arg (forming hydrogen bonds to the base pairs)²¹. The conserved stretch of amino acids following His 23 suggests that each repeat represents an extended DNA-binding finger domain linked by invariant joints. Each finger possibly interacts with about six nucleotides—half a DNA period (for details see ref. 18).

This structural feature of the *Kr* protein makes it different from the other segmentation and homoeotic gene products so far analysed. More importantly, it demonstrates a connection between a segmentation gene and a transcription factor known to bind to DNA (see Fig. 3), which suggests that DNA binding may be an important feature of *Kr* function, as proposed for other segmentation and homoeotic genes of *Drosophila* which contain the conserved 'homoeo-box' domain^{15,16,22–27}. However, in view of the unrelated protein structure, the interaction of the *Kr* protein with DNA should be different from that of the homoeo-box domain proteins¹⁵. Whereas the latter domain provides general DNA-binding features that may become specific through variability in the adjacent sequences, as found in all genes so far analysed^{15,16,22–27}, the *Kr* finger motif could possibly specify directly for the DNA-binding site. In analogy to TFIIIA and 5S RNA regulation during *Xenopus* development (see refs 19, 20 for details), the *Kr* finger domain could then bind to specific DNA sequences to form stable transcriptional com-

plexes which act on few other (sets of) specific genes, and it might also be involved in autoregulation as outlined in Meinhardt's model of pattern formation^{28,29}. In addition, the *Kr* protein may bind to RNA (see below).

In looking for possible DNA-binding sites on the *Kr* gene, it was interesting to note a six-nucleotide repeat (ACAAAA) (see Fig. 4) that could be a target for maternal factors which are expected to initiate and regulate *Kr* gene expression^{11,13} and/or a target for the *Kr* protein for autoregulation processes (see above). The 'ACAAAA box' is repeated eight times (on both DNA strands and in both orientations) within 180 bp downstream from the *Kr* TATA box and five times downstream from the TGA stop codon (Fig. 4a), forming part of the non-translated sequences. This hexanucleotide repeat and its location within the transcribed region of the *Kr* gene are reminiscent of the multiple contact binding of other transcription factors³⁰, including a single finger loop of TFIIIA¹⁸, thus, as a multiple-finger protein, *Kr* (in analogy to TFIIIA) may bind many such recognition sites within the transcribed region of genes. As there is no need for multiple fingers merely to bind DNA, it would be reasonable to speculate that the targets for the *Kr* protein binding might be internal (see the above ACAAAA box) and/or that its multiple fingers allow binding during the rapid DNA replication that occurs at the time *Kr* is first expressed^{12,13}. In addition, as the ACAAAA box occurs three times in a hypothetical secondary structure of *Kr* mRNA (Fig. 4b) and twice in mRNA of the pair-rule segmentation gene³¹ *fushi tarazu* (*ftz*; Fig. 4c), it could allow the *Kr* protein to interact with its own gene and/or the derived mRNA (for autoregulation at the transcriptional and/or translational level) and to regulate *ftz* gene expression in the *Kr* domain of the embryo defined by *Kr* mutants. The latter proposal is based on changes in the *ftz* expression pattern in *Kr*⁻ embryos which suggest that *Kr*⁺ activity is required for normal *ftz* gene expression (S. Carroll and M. Scott; P. Ingham; W. J. Gehring, personal communications).

In *Drosophila*, the finger motif is seen in several proteins encoded by the *serendipity* (*sry*) locus^{32,33}. This observation and the structural homology to the frog TFIIIA suggest conservation of the finger motif. We suspect an ancestral single functional unit that binds to a half-turn of DNA¹⁸. During evolution, the finger unit may have multiplied by sequence duplication or gene conversion³⁴ and adjusted to more subtle and specialized sequences by joining several similar units in series. In this way, the many-fingered protein offers an ideal way of recognizing specific DNA sequences through changing the amino acids at the fingertips and, therefore, modulating the DNA binding. This allows a huge number of combinatorial possibilities for specific DNA binding, all based on this structural scheme. This combinatorial aspect resembles protein recognition and protein binding by antibodies, and in view of this, it is not surprising

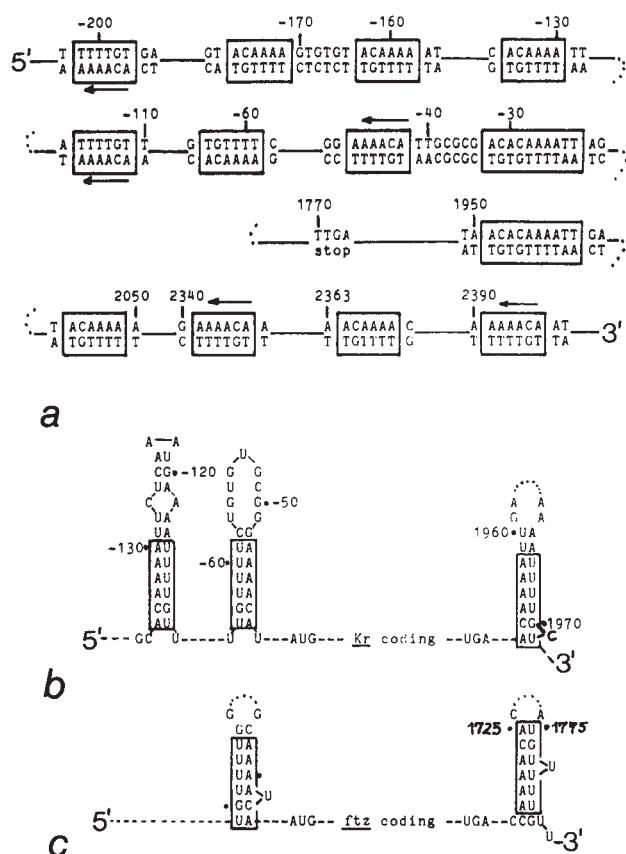


Fig. 4 Distribution of small repeated sequences in *Kr* genomic DNA and in a presumptive secondary structure of *Kr* and *ftz* mRNA. *a*, Nucleotide sequence of *Kr* untranslated 5' and 3' sequences. Nucleotide numbers refer to positions in Fig. 2. Note the eight repeats of the AAAAAA sequence (orientation indicated by arrows) within 180 bp at the 5' end, the five repeats at the 3' end and an extended 10-bp repeat (large boxes) at both the 5' and 3' end. A hypothetical palindrome structure of *Kr* (*b*) and *ftz* (*c*) mRNA leaving two and one AAAAAA palindrome at the 5' and 3' end of *Kr* mRNA, respectively, and one each at the 5' and 3' end of *ftz* mRNA. Note that the 5' palindrome was recovered in *ftz* cDNA (for details and discussion see ref. 15, Fig. 4) and in *Kr* pcK2b cDNA (see Fig. 2), being part of the hairpin structure which possibly serves as primer for the synthesis of the second cDNA strand. The numbering of nucleotides refers to Fig. 2.

to detect, under conditions of reduced stringency, *Kr*-homologous sequences (S.C. *et al.*, in preparation) which may reflect *Kr*-related members of the 'finger protein gene family' in addition to TFIIA, *sry* and *Kr*.

We thank M. Rosbash and A. Vincent for pointing out the Klug paper to us and for sharing unpublished data, and M. Goldschmidt-Clermont, David Marcey, P. Ekblom, Ch. Nüsslein-Volhard, A. Gierer, R. Lehmann, H. Saumweber, U. Schwarz, D. Marcey, T. Berleth, U. Nauber, I. Baxivanelis and R. Groemke-Lutz for their many and various contributions. S.C. was supported by a NSERC of Canada postdoctoral fellowship. *Note added in proof:* The TFIIA structure has also been described by Brown *et al.*³⁵.

Received 9 October; accepted 19 November 1985.

1. Wieschaus, E. & Gehring, W. *Devl Biol.* **50**, 249-263 (1976).
2. Szabad, J., Schüpbach, T. & Wieschaus, E. *Devl Biol.* **73**, 265-271 (1979).
3. Simcox, A. A. & Sang, J. H. *Devl Biol.* **97**, 212-221 (1983).
4. Nüsslein-Volhard, C. in *Determinants of Spatial Organisation* (eds Subtelney, S. & Konigsberg, I. R.) 185-211 (Academic, New York, 1979).
5. Schubiger, G., Moseley, R. C. & Wood, W. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2050-2053 (1977).

6. Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 283-295 (1980).
7. Wieschaus, E., Nüsslein-Volhard, C. & Jürgens, G. *Wilhelm Roux Arch. dev. Biol.* **193**, 296-307 (1984).
8. Nüsslein-Volhard, C., Wieschaus, E. & Kluding, H. *Wilhelm Roux Arch. dev. Biol.* **193**, 267-282 (1984).
9. Jürgens, G., Kluding, H., Nüsslein-Volhard, C. & Wieschaus, E. *Wilhelm Roux Arch. dev. Biol.* **193**, 283-295 (1984).
10. Nüsslein-Volhard, C., Wieschaus, E. & Jürgens, G. *Verh. dt. zool. Ges.*, 91-104 (1982).
11. Wieschaus, E., Nüsslein-Volhard, C. & Kluding, G. *Devl Biol.* **104**, 172-186 (1984).
12. Preiss, A., Rosenberg, U. B., Kienlin, A., Seifert, E. & Jäckle, H. *Nature* **313**, 27-32 (1985).
13. Knipple, D. C. *et al. Nature* **317**, 40-44 (1985).
14. Kozak, M. *Nucleic Acids Res.* **12**, 857-872 (1984).
15. Laughton, A. & Scott, M. P. *Nature* **310**, 25-31 (1984).
16. Poole, S., Kauvar, L. M., Drees, B. & Kornberg, T. *Cell* **40**, 37-43 (1985).
17. Marshall, R. D. *et al. Rev. Biochem.* **41**, 673-702 (1972).
18. Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **4**, 1609-1614 (1985).
19. Pelham, H. R. B. & Brown, D. D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4170-4174 (1980).
20. Hoda, B. W. & Roeder, R. G. *Cell* **22**, 119-126 (1980).
21. Ohlendorf, D. H. & Matthews, B. W. *Rev. Biophys. Bioengng* **12**, 259-284 (1983).
22. McGinnis, W., Levine, M., Hafen, E., Kuroiwa, A. & Gehring, W. J. *Nature* **308**, 428-433 (1984).
23. McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. *Cell* **37**, 403-408 (1984).
24. Carrasco, A. E., McGinnis, W., Gehring, W. J. & De Robertis, E. M. *Cell* **37**, 409-414 (1984).
25. Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., De Robertis, E. M. & Gehring, W. J. *Nature* **310**, 70-71 (1984).
26. Levine, M., Rubin, G. & Tjian, R. *Cell* **38**, 667-673 (1984).
27. Fjose, A., McGinnis, W. J. & Gehring, W. J. *Nature* **313**, 284-289 (1985).
28. Meinhard, H. *Ber. Bunsenges. phys. Chem.* **89**, 691-699 (1985).
29. Meinhard, H. *J. Cell Sci.* (in the press).
30. Dyan, W. S. & Tjian, R. *Nature* **316**, 774-777 (1985).
31. Kuroiwa, A., Hafen, E. & Gehring, W. J. *Cell* **37**, 825-831 (1984).
32. Vincent, A., O'Connell, P., Gray, M. R. & Rosbash, M. *EMBO J.* **3**, 1003-1013 (1984).
33. Vincent, A., Colot, H. V. & Rosbash, M. *J. molec. Biol.* **186**, 149-166 (1985).
34. Hood, L. M., Campbell, J. H. & Elgin, S. C. R. *A. Rev. Genet.* **9**, 305-334 (1975).
35. Brown, R. S., Sander, C. & Argos, P. *FEBS Lett.* **186**, 271-274 (1985).

Induction of yeast Ty element transcription by ultraviolet light

Mark Rolfe, Ad Spanos & Geoffrey Banks

Genetics Division, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Gene expression induced in response to DNA damage is now well characterized in *Escherichia coli* (reviewed in ref. 1), and the functions of the proteins encoded by some of these genes are also understood. Physiological evidence has implicated similar responses in eukaryotes²⁻⁹, and recently DNA sequences whose transcription can be induced by ultraviolet (UV) irradiation have been identified and cloned from the yeast *Saccharomyces cerevisiae*¹⁰⁻¹². With the exception of the *Ustilago maydis* Recl protein, however¹³, no inducible eukaryotic proteins have been characterized with respect to their role in post-irradiation DNA metabolism. Using differential colony hybridization methods, we have described the cloning of four *S. cerevisiae* DNA sequences whose transcript levels increase ~10-fold after a dose of UV irradiation which was previously shown to induce two novel proteins in yeast^{12,14}. Although they possess different restriction enzyme maps, these sequences did share elements of homology as defined by cross-hybridization experiments. We show here that all four sequences possess homology with yeast Ty elements. These elements comprise a family of heterogeneous, dispersed and moderately repetitive transposable DNA sequences which are present as ~30 copies per haploid genome in laboratory strains and whose archetypal structure is shown in Fig. 1 (refs 15, 16 and reviewed in ref. 17). We also show directly that the 5.7-kilobase (kb) major Ty transcript, initiated in one terminal direct repeat sequence (δ) and terminated in the other¹⁸, is induced by UV irradiation.

To confirm that the cloned yeast DNA in plasmids pMR1, pUV1, pUV2 and pUV3 (ref. 12) contained Ty element sequences, chromosomal DNA was cleaved by the *Xho*I restriction enzyme and hybridized to each labelled plasmid and to an intact Ty1-15 element in turn. Most Ty elements contain a *Xho*I cleavage site in each δ sequence, but none in the intervening sequences, and so genomic cleavage generates characteristic 5.7-kb Ty fragments (see Fig. 1). Figure 2 shows that each plasmid hybridized to such a 5.7-kb *Xho*I chromosomal DNA

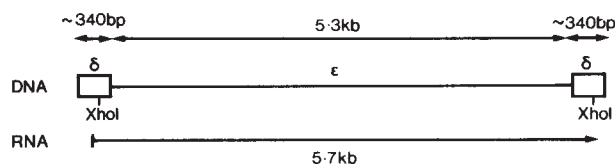


Fig. 1 Generalized Ty element structure. Delta (δ) sequences of ~ 0.34 kb, usually containing *XhoI* restriction enzyme sites, are separated by ~ 5.3 kb of epsilon (ϵ) sequence. Transcription is initiated in one δ , extends through the ϵ sequence and terminates in the other δ , producing a 5.7-kb RNA transcript^{15,16,18}.

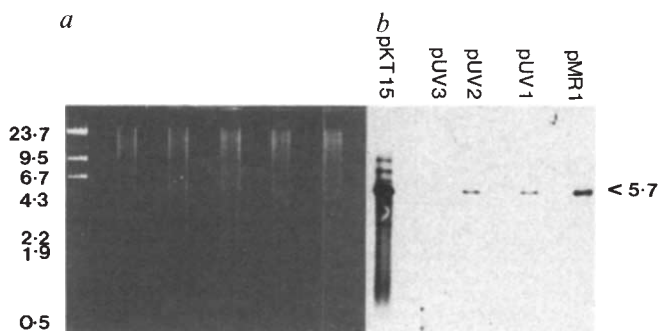


Fig. 2 Hybridization of chromosomal Ty elements to the four cloned inducible yeast DNA sequences¹². Total *S. cerevisiae* YNN27 DNA was digested with *XhoI* enzyme, electrophoresed and transferred to nitrocellulose²⁵. Plasmids pKT15, pMR1, pUV1, pUV2 and pUV3 were labelled by nick-translation²⁶ and hybridized to the immobilized *XhoI* fragments²⁷. The 5.7-kb *XhoI* fragment characteristic of Ty elements is arrowed. *HindIII*-digested λ DNA was run as a size marker (left-hand lane, with fragment sizes in kb). *a*, Ethidium bromide fluorescence of DNA digests after electrophoresis. *b*, Hybridizations to the DNA probes (shown above the lanes) after transfer of the digests to nitrocellulose.

fragment, indistinguishable from that to which the Ty1-15 sequence hybridized. Specific plasmid to plasmid hybridization experiments with Ty and δ probes also confirmed this conclusion (data not shown).

We have previously shown that sequences in all four of the cloned yeast DNA fragments hybridize to yeast transcripts of identical size whose abundances increase ~ 10 -fold after UV irradiation¹². To identify these transcripts more definitively, equal amounts of total RNA from UV-irradiated and from unirradiated yeast cells were probed with labelled pMR1 and with pKT1-15 (containing the Ty1-15 element) DNA after electrophoretic separation and Northern transfer of the RNA. Figure 3*a* shows that both probe sequences hybridized to transcripts of identical size (5.7 kb) whose abundance after UV irradiation increased about 10-fold. Hybridization of these RNAs with a labelled plasmid encoding the non-inducible yeast proteins H2A and Protein 1 demonstrated the specificity of the induction of the Ty element transcription (Fig. 3*b*).

The evidence presented above demonstrates that transcription of yeast transposable Ty elements is strongly stimulated by UV irradiation and identifies the δ sequence promoter as being UV-inducible. Because our experiments measured steady-state messenger RNA levels, it is possible that an increase in mRNA stability contributes to the observed increase in transcript levels after UV irradiation. It has also been established recently that the transcription of the *Drosophila melanogaster copia* element is induced by environmental stress such as elevated temperature and exposure to certain chemicals¹⁹. The elegant experiments of Boeke and co-workers have established that increased levels of Ty transcripts are correlated with increased transposition of marked Ty elements to a target plasmid and to chromosomal locations²⁰. Their experiments also implicate an RNA intermediate in the pathway of transposition, thereby inviting comparison with the structure and mode of integration of retroviruses into

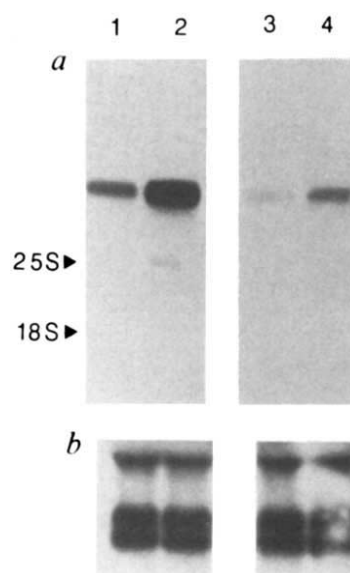


Fig. 3 *a*, Hybridization of pKT15 (lanes 1, 2) and pMR1 (lanes 3, 4) DNAs to RNA from unirradiated (lanes 1, 3) or UV-irradiated (lanes 2, 4) *S. cerevisiae* YNN27 cells. The cells were irradiated and total RNA extracted as described previously¹². The resulting RNA was denatured by glyoxal and dimethyl sulphoxide, electrophoresed through a 1% agarose gel²⁸ and transferred to Gene Screen Plus membranes before hybridization. The mobilities of the two largest ribosomal RNAs are indicated. *b*, Control hybridization of the RNAs with a plasmid pMH501 probe. pMH501 contains an *SstI* fragment of TRT-1 DNA in the vector pEMBL18 and thus carries the yeast H2A and Protein 1 genes²⁹. The upper band is an uncharacterized, strain-dependent hybridization (L. Capsey, personal communication).

a host chromosome (refs 20–22 and reviewed in refs 23, 24). We are now investigating whether the increased Ty element transcription induced by UV irradiation of yeast cells is accompanied by Ty transposition events. Such a process may well contribute to the induction of mutations and other chromosomal aberrations by UV irradiation.

We thank Drs Alan Kingsman, Peter Phillipsen and Ron Davis for gifts of plasmids used in this work, and Steven Sedgwick and Robin Holliday for discussions.

Received 16 May; accepted 3 December 1985.

- Walker, G. C. *Microbiol. Rev.* **48**, 60–93 (1984).
- Holliday, R. *Mutat. Res.* **29**, 149–153 (1975).
- Moore, P. D. *Mutat. Res.* **28**, 367–380 (1975).
- Fabre, F. & Roman, H. L. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1667–1671 (1977).
- Leaper, S., Resnick, M. A. & Holliday, R. *Genet. Res. Camb.* **35**, 291–307 (1980).
- Stadler, D. & Moyer, R. *Genetics* **98**, 763–774 (1981).
- Lee, M. G. & Yarranton, G. T. *Molec. gen. Genet.* **185**, 245–250 (1982).
- Baker, T. I. *Molec. gen. Genet.* **190**, 295–299 (1983).
- Siede, W., Eckardt, F. & Brendel, M. *Molec. gen. Genet.* **190**, 406–412 (1983).
- McClanahan, T. & McEntee, K. *Molec. cell. Biol.* **4**, 2356–2363 (1984).
- Ruby, S. W. & Szostak, J. W. *Molec. cell. Biol.* **5**, 75–84 (1985).
- Rolfe, M. *Curr. Genet.* **9**, 533–538 (1985).
- Kmiec, E., Brougham, M., Yarnall, M., Kroeger, P. E. & Holloman, W. K. *UCLA Symp. molec. cell. Biol.* **10**, 753–760 (1983).
- Rolfe, M. *Curr. Genet.* **9**, 529–532 (1985).
- Cameron, J. R., Loh, E. Y. & Davis, R. *Cell* **16**, 739–751 (1979).
- Kingsman, A. J., Gimlich, R. L., Clarke, L., Chinault, A. C. & Carbon, J. J. *molec. Biol.* **145**, 619–632 (1981).
- Williamson, V. M. *Int. Rev. Cytol.* **83**, 1–25 (1983).
- Elder, R. T., St John, T. P., Stinchcomb, D. T. & Davis, R. W. *Cold Spring Harb. Symp. quant. Biol.* **45**, 581–584 (1981).
- Strand, D. J. & McDonald, J. F. *Nucleic Acids Res.* **13**, 4401–4410 (1985).
- Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. *Cell* **40**, 491–500 (1985).
- Mellor, J. *et al. Nature* **313**, 243–246 (1985).
- Hauber, T., Nelbock-Hochstatter, P. & Feldmann, H. *Nucleic Acids Res.* **13**, 2745–2758 (1985).
- Baltimore, D. *Cell* **40**, 481–482 (1985).
- Varmus, H. E. *Nature* **314**, 583–584 (1985).
- Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. J. *molec. Biol.* **113**, 737–751 (1977).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning*, 124 (Cold Spring Harbor Laboratory, New York, 1982).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201–5205 (1980).
- Hereford, L., Fahrner, K., Woolford, J. Jr, Rosbash, M. & Kaback, D. B. *Cell* **18**, 1261–1271 (1979).

Physics

Aspects of Matter in Science Today: Discussion with Emile Noël. W.J. DUFFIN (ed.). W.J. Duffin, 14 Orchard Croft, Cottingham, North Humberside HU16 4HG, England: 1985. Pp.196. Pbk ISBN 0-9510438-0-3. Pbk £3.95.

Classical and Quantum Field Theory of Exactly Soluble Nonlinear Systems. By PIOTR GARBA-CZEWSKI. Wiley: 1985. Pp.244. ISBN 9971-966-55-7. £33.85.

Collisions of Fast Charged Particles in Solids. By N.P. KALASHNIKOV, V.S. REMIZOVICH and M.I. RYAZANOV. KONSTANTIN ERASOV (trans.). Gordon & Breach: 1985. Pp.438. Hbk ISBN 2-88124-039-9. \$145.00

Einstein's Relativity The Greatest Fallacy in the Twentieth Century. By SHARAD TIPNIS. Madhav Publications, 40/1 Dattakrupa, Adarshnagar, Pune-Satara Road, Pune-411 037, India: 1985. Pp.199. Rs.90.

Electroweak Effects at High Energies. HARVEY B. NEWMAN (ed.). Plenum: 1985. Pp.866. ISBN 0-306-41904-1. Np.

Entropy, Large Deviations, and Statistical Mechanics. By RICHARD S. ELLIS. Springer-Verlag: 1985. Pp.364. Hbk ISBN 3-540-96052-X. DM 184.

Fourier Transforms in Physics. By D.C. CHAM-PENEY. Adam Hilger: 1985. Pp.64. Pbk ISBN 0-85274-794-2. Pbk £2.95.

Fundamentals of Laser Interactions. F. EHLOTZKY (ed.). Springer-Verlag: 1985. Pp.314. Pbk ISBN 3-540-15640-2/0-387-15640-2. Np.

Large-scale Computations in Fluid Mechanics. BJORN E. ENGOQUIST, STANLEY OSHER and RICHARD C.J. SOMERVILLE (eds). American Mathematical Society, Providence, Rhode Island: 1985. Pp.370. Hbk ISBN 0-8218-1122-3. Np.

Lecture Notes in Physics. Hadrons and Heavy Ions, Proceedings, Cape Town, South Africa 1984. W.D. HEISS (ed.). Springer-Verlag: 1985. Pp.458. Pbk ISBN 3-540-15653-4. DM 73.

Macroscopic Modelling of Turbulent Flows. U. FRISCH et al. (eds). Springer-Verlag: 1985. Pp.360. Pbk ISBN 3-540-15644-5/0-387-15644-5. Pbk DM 45.

Magnetospheric and Ionospheric Plasmas. E.R. SCHMERLING, S.W.H. COWLEY and P.H. REIFF (eds). Pergamon: 1985. Pp.426. Pbk ISBN 0-08-033193-9. Pbk \$49.50, £31.

Maxwell's Equations and their Applications. By E.G. THOMAS and A.J. MEADOWS. Adam Hilger: 1985. Pp.54. Pbk ISBN 0-85274-778-0. Pbk £2.95.

Mechanics of Continua and Wave Dynamics. By L. BRETHOVSKIKH and V. GONCHAROV. Springer-Verlag: 1985. Pp.342. ISBN 3-540-13765-3. DM 139.

Muon Spin Rotation Spectroscopy: Principles and Applications in Solid State Physics. By A. SCHENCK. Adam Hilger: 1985. Pp.325. ISBN 0-85274-551-6. £39, \$58.

Oscillations and Waves. By R. BUCKLEY. Adam Hilger: 1985. Pp.64. Pbk ISBN 0-85274-793-4. Pbk £2.95.

Physics of Magnetic Materials, Jadwisin '84 (Poland). J. RAULUSZKIEWICZ, H. SZYMCAK and H.K. LACHOWICZ (eds). Wiley: 1985. Pp.580. ISBN 9971-978-34-2. £57.15.

Quantum Reality: Beyond the New Physics. By NICK HERBERT. Anchor/Doubleday: 1985. Pp.268. ISBN 0-385-18704-1. \$16.95.

Quarks, Leptons, and Beyond. H. FRITZSCH, R.D. PEECEI, H. SALLER and F. WAGNER (eds). Plenum: 1985. Pp.555. ISBN 0-306-41925-4. Np.

Roundabout: The Physics of Rotation in the Everyday World. Readings from Scientific American. By JEARL WALKER. W.H. Freeman: 1985. Pp.70. Hbk ISBN 0-7167-1724-7; pbk ISBN 0-7167-1725-5. Hbk \$19.95; pbk \$11.95.

Spectrometric Techniques, Vol.4. GEORGE A. VANASSE (ed.). Academic: 1985. Pp.262. Hbk ISBN 0-12-710404-6. \$75.50, £65.50.

Springer Series in Chemical Physics, Vol.41: High Resolution Spectral Atlas of Nitrogen Dioxide 559 - 597 nm. By K. UEHARA and H. SASADA. Springer-Verlag: 1985. Pp.226. ISBN 3-540-15027-7, 0-387-15027-7. DM 99.

Biological Sciences

The Grizzlies of Mount McKinley. By ADOLPH MURIE. University of Washington Press: 1985. Pp.270. Pbk ISBN 0-295-96204-6. Pbk \$9.95.

Growth, Health and Fitness of Teenager. H.C.G. KEMPER (ed.). Karger: 1985. Pp.202. ISBN 3-8055-4042-6. DM 139, \$49.50.

Habitat Selection in Birds. MARTIN L. CODY (ed.). Academic: 1985. Pp.558. ISBN 0-12-178080-5. \$69.50, £69.50.

Hawk-Moths of the British Isles. By MICHAEL EASTERBROOK. Shire Natural History: 1985. Pp.24. Pbk ISBN 0-85263-743-8. Pbk £1.25.

Hormones: The Messengers of Life. By LAWRENCE CRAPO. Freeman: 1985. Pp.194. Hbk ISBN 0-7167-1757-3; pbk ISBN 0-7167-1753-0. Hbk \$19.95; pbk \$11.95.

Human Nature: Darwin's View. By ALEXANDER ALLAND JR. Columbia University Press: 1985. Pp.242. Hbk ISBN 0-231-05898-5. \$25.

Immune Recognition of Protein Antigens. Current Communications in Molecular Biology. W. GRAEME LAVER and GILLIAN M. AIR. Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, New York 11724, USA: 1985. Pp.197. Pbk ISBN 87969-185-9. \$30.

Immunology of the Bacterial Cell Envelope. D.E.S. STEWART-TULL and M. DAVIES (eds). Wiley: 1985. Pp.391. ISBN 0-471-90551-6. £40.

Interface between Neurology and Psychiatry. M.R. TRIMBLE (ed.). Karger: 1985. Pp.190. ISBN 3-8055-4023-X. DM 137, \$48.50.

International Review of Cytology. G.H. BOURNE and J.F. DANIELLI (eds). Academic: 1985. Pp.306. ISBN 0-12-364494-1. \$49.50, £49.50.

International Review of Neurobiology, Vol. 26. JOHN R. SMYTHIES and RONALD J. BRADLEY (eds). Academic: 1985. Pp.342. ISBN 0-12-366826-3. \$65, £65.

Intrinsic Motivation and Self-Determination in Human Behaviour. By EDWARD L. DECI and RICHARD M. RYAN. Plenum: 1985. Pp.371. Hbk ISBN 0-306-42022-8. Np.

Mamíferos Silvestres de la Cuenca de Mexico. By GERARDO CEBALLOS GONZALEZ and CARLOS GALINDO LEAL. Instituto de Ecología, A.C., Apartado Postal 18-845, Deleg. Miguel Hidalgo, 11800 Mexico, DF: 1985. Pp.299. ISBN 968-18-1700-1. \$15.

Mechanisms of Cell-Mediated Cytotoxicity II. PIERRE HENKART and ERIC MARTZ (eds). Plenum: 1985. Pp.568. ISBN 0-306-42009-0. Np.

Mechanisms of Cytotoxicity by NK Cells. RONALD B. HERBERMAN and D.M. CALLEWAERT (eds). Academic: 1985. Pp.669. Hbk 0-12-341370-2. £49.50, \$57.

Modification of Proteins During Aging. RICHARD C. ADELMAN and EUGENE E. DEKKER (eds). Alan R. Liss: 1985. Pp.122. Hbk ISBN 0-8451-2307-6. £25.

The Modulation of Immunity. M.S. MITCHELL (ed.). Pergamon: 1985. Pp.478. ISBN 0-08-031977-7. £80, \$130.

Molecular Basis of Nerve Activity. J.-P. CHANGEUX, F. HUCHO, A. MAELICKE and E. NEUMANN (eds). Walter de Gruyter & Co., 13 Genthiner Str., 1000 Berlin 30, FRG: 1985. Pp.784. Hbk ISBN 311-010345-1. DM 260.

The Molecular Biology of the Bacilli. DAVID A. DUBNAU (ed.). Academic: 1985. Pp.259. ISBN 0-12-222770-6. \$49, £49.

Molecular Structure of the Number 21 Chromosome and Down's Syndrome. GEORGE F. SMITH (ed.). New York Academy of Sciences: 1985. Pp.250. Pbk ISBN 0-89766-295-4. Np.

Molecule, Nerve and Embryo. By RICHARD R. RIBCHESTER. Blackie: 1985. Pp.204. Hbk ISBN 0-216-91800-6. £14.95.

Neuroendocrine Perspectives. EUGENIO E. MULLER, ROBERT U.M. MACLEOD and LAWRENCE A. FROHMAN (eds). Elsevier Biomedical Press: 1985. Pp.301. Hbk ISBN 0-444-80665-2. Dfl 216, \$65.

Neuro-Oncology. F. CLIFFORD ROSE and W.S. FIELDS (eds). Karger: 1985. Pp.274. Hbk ISBN 3-8055-4050-X. SwFr. 230, DM 276, £71.90, \$98.

Organized Multienzyme Systems Catalytic Properties. G. RICKEY WELCH (ed.). Academic: 1985. Pp.458. Hbk ISBN 0-12-744040-2. £65, \$75.

The Origin and Evolution of Sex. MBL Lectures in Biology, Vol. 7. HARLYN O. HALVORSON and ALBERTO MONROY (eds). Alan R. Liss: 1985. Pp.350. Hbk ISBN 0-8451-2206-1. £52.

Oro-Facial Pain and Neuromuscular Dysfunction: Mechanisms and Clinical Correlates. I. KLINBERG and B. SESSLE (eds). Pergamon: 1985. Pp.229. ISBN 0-08-032001-5. £48, \$82.

Parasitic Worms. By JIM FLEGG. Shire Natural History: 1985. Pp.24. Pbk ISBN 0-85263-761-6. Pbk £1.25.

The Peccaries. By LYLE K. SOWLS. University of Arizona Press: 1984. Pp.251. ISBN 0-8165-0822-4. \$22.50.

Pelargonium Family. By WILLIAM J. WEBB. Croom Helm: 1984. Pp.104. ISBN 0-7099-2737-7. £19.95.

Physiological Strategies in Avian Biology. By J.G. PHILLIPS, P.J. BUTLER and P.J. SHARP. Blackie: 1985. Pp.218. Hbk ISBN 0-412-00921-8; pbk ISBN 0-412-00931-5. Hbk £19.95; pbk £9.95.

The Pineal Gland: Endocrine Aspects. G.M. BROWN and S.D. WAINWRIGHT (eds). Pergamon: 1985. Pp.367. Pbk ISBN 0-08-031991-2. Pbk £31.

Protein and Nucleic Acid Structure and Dynamics. JONATHAN KING (ed.). Addison-Wesley, Wokingham, Berkshire, UK: 1985. Pp.587. Hbk ISBN 0-8053-5403-4. £29.95.

Proteinuria. M.M. AVRAM (ed.). Plenum: 1985. Pp.232. ISBN 0-306-41956-4. Np.

The Puffin. By JIM FLEGG. Shire Natural History: 1985. Pp.24. Pbk ISBN 0-85263-744-6. Pbk £1.25.

The Pulmonary Circulation and Acute Lung Injury. SAMI I. SAID (ed.). Futural Publishing, 295 Main Street, Mount Kisco, New York 10549: 1985. Pp.472. ISBN 0-87993-249-X. \$65.

The Receptor, Vol. 2. P. MICHAEL CONN (ed.). Academic: 1985. Pp.436. Hbk ISBN 0-12-185202-4. £56.50, \$65.

Regulation of Pituitary Function. Tj.B. VAN WIMERSMA GREIDANUS and S.W.J. LAMBERTS (eds). Karger: 1985. Pp.218. Hbk ISBN 3-8055-4061-2. SwFr. 179, DM 214, £56, \$76.25.

Sharks of the World. By RODNEY STEEL. Blandford Press, Link House, West Street, Poole, Dorset BH15 1LL, UK: 1985. Pp.192. Hbk ISBN 0-7137-1425-5. £10.95.

Sleep: Neurotransmitters and Neuromodulators. A. WAUGUIER et al. (eds). Raven: 1985. Pp.344. ISBN 0-88167-102-9. \$63.

Snail. By JENS OLESEN. A. and C. Black: 1985. Pp.25. Hbk ISBN 0-7136-2708-5. £3.95.

Spectroscopy and the Dynamics of Molecular Biological Systems. PETER M. BAYLEY and ROBERT E. DALE (eds). Academic: 1985. Pp.406. ISBN 0-12-083240-2. \$42.50, £38.50.

Spectroscopy of Biological Molecules. A.J.P. ALIX, L. BERNARD and M. MANFAIT (eds). Wiley: 1985. Pp.465. ISBN 0-471-90883-5. £28.

The Starling. By C.J. FEARE. Shire Natural History: 1985. Pp.24. Pbk ISBN 0-85263-764-0. Pbk £1.25.

Statistics in Ornithology. B.J.T. MORGAN and P.M. NORTH (eds). Springer-Verlag: 1985. Pp.418. Pbk ISBN 3-540-96189-5. DM 65.

The Status of Ecological Reserves in Canada. By P.M. TASCHEREAU. Institute for Resource and Environmental Studies: 1985. Pp.120. Pbk ISBN 0-7703-0462-1. Pbk \$8.

Structure and Function of Sarcoplasmic Reticulum. SIDNEY FLEISCHER and YUJI TONOMURA (eds). Academic: 1985. Pp.597. ISBN 0-12-260380-X. \$49, £49.

Tidal Flat Ecology: An Experimental Approach to Species Interactions. By KARSTEN REISE. Springer-Verlag: 1985. Pp.191. Hbk ISBN 3-540-15447-7. DM 98.

Viral Haemorrhagic Fevers: Report of a WHO Expert Committee. Technical Report Series 721. World Health Organization: 1985. Pp.126. Pbk ISBN 92-4-120721-3. SwFr. 10.

Whales, Dolphins and Seals and their Kin. Longman Nature Library, Longman:1985. Pp.130. Hbk ISBN 0-582-89262-7. Np.

Willows of the British Isles. By THERESA BRENDELL. Shire Natural History:1985. Pp.24. Pbk ISBN 0-85263-765-9. Pbk £1.25.

The Wolves of Mount McKinley. By ADOLPH MURIE. University of Washington Press:1985. Pp.258. Pbk ISBN 0-295-96203-8. Pbk \$9.95.

The Year in Immunology 1984-85. J.M. CRUSE and R.E. LEWIS, Jr. (eds). Karger:1985. Pp.234. ISBN 3-8055-4025-6. DM 189, \$67.25.

Yeast Identification Program. By J.A. BARNETT, R.W. PAYNE and D. YARROW. Cambridge Micro Software (Cambridge University Press):1985. For DEC Rainbow using MS-DOS. ISBN 0-521-30918-2. £85.

Applied Biological Sciences

Advances in Enzyme Regulation 23. By GEORGE WEBER. Pergamon:1985. Pp.463. ISBN 0-08-032727-3. £99, \$160.

Advances in Parasitology. Vol. 23. J.R. BAKER and R. MULLER (eds). Academic:1984. Pp.255. ISBN 0-12-031723-0. \$40, £27.

Animal Cell Biotechnology, Vol. 2. R.E. SPIER and J.B. GRIFFITHS (eds). Academic:1985. Pp.475. Hbk ISBN 0-12-657552-5. £59, \$66.

Approaches to Antiviral Agents. MICHAEL R. HARNDEN (ed.). Macmillan Press:1985. Pp.350. Hbk ISBN 0-333-35928-3. £45.

Autoimmunity and Endocrine Disease. ROBERT VOLPÉ (ed.). Marcel Dekker:1985. Pp.488. ISBN 0-8247-7311-X. \$75.

Bacterial Endotoxins: Structure, Biomedical Significance, and Detection with the Limulus Amebocyte Lyse Test. JAN W. ten CATE, HARRY R. BULLER, AUGUSTE STURK and JACK LEVIN (eds). Alan R. Liss:1985. Pp.466. Hbk ISBN 0-8451-5039-1. £52.

Biochemical Pharmacology and Toxicology. Vol. 1, Methodological Aspects of Drug Metabolizing Enzymes. D. ZAKIM and D.A. VESSEY (eds). Wiley:1985. Pp.372. Hbk ISBN 0-471-86793-4. £66.50.

Bone Resorption, Metastasis, and Diphosphonates. SILVIO GARATTINI (ed.). Raven Press:1985. Pp.221. Hbk ISBN 0-88167-137-1. \$43.50.

Burkitt's Lymphoma: A Human Cancer Model. G. LENOIR, G. O'CONNOR and C.L.M. OLWENY (eds). International Agency for Research on Cancer, Lyon:1985. Pp.484. Hbk ISBN 0-19-723057-1. £20.

Carcinoma of the Large Bowel and its Precursors. J.R.F. INGALL and A.J. MASTROMARINA (eds). Alan R. Liss:1985. Pp.305. ISBN 0-8451-5036-7. £37.

Cardiovascular Shunts: Phylogenetic, Ontogenetic and Clinical Aspects. KJELL JOHANSEN and WARREN BURGREN (eds). Munksgaard:1985. Pp.492. ISBN 87-16-09822-6. D. Kr.350.

Care of the Burn Wound: Wound Healing, Grafts and Transplantation, Synthetic Skin Substitutes. S. RANDOLPH MAY and GIOVANNI DOGO (eds). Karger:1985. Pp.246. ISBN 3-8055-3991-6. DM264, \$93.75.

Chemically Induced Birth Defects. By JAMES L. SCHARDEIN. Dekker:1985. Pp.904. ISBN 0-8247-7247-4. \$125.

Chemotherapy of Gastro-Intestinal Helminths. H. VANDEN BOSSCHE, D. THIENPONT and P.G. JANSSENS (eds). Springer-Verlag:1985. Pp.719. ISBN 3-540-13111-6/0-387-13111-6. DM660.

Clinical Oncology and Cancer Nursing. F.J. CLETON et al. (eds). Pergamon:1985. Pp.125. Pbk ISBN 0-08-030786-8. Pbk £21, \$34.

Cooperative Approaches to Research and Development of Orphan Drugs. MELVIN H. VAN WOERT and EUNYONG CHUNG (eds). Alan R. Liss:1985. Pp.210. Hbk ISBN 0-8451-5047-2. \$36.

The Cytobiology of Leukaemias and Lymphomas. D. QUAGLINO and F.G.J. HAYHOE (eds). Raven Press:1985. Pp.524. Hbk ISBN 0-88167-091-X. \$65.50.

Diagnosis and Treatment of Erectile Disturbances: A Guide for Clinicians. R. TAYLOR SEGRAVES and HARRY W. SHEONBERG (eds). Plenum:1985. Pp.253. ISBN 0-306-41871-1. Np.

Diagnostic Accuracy: Effect on Treatment Planning. ROBERT J. SHPRINTZEN and NATALIE W. PAUL (eds). Alan R. Liss:1985. Pp.151. ISBN 0-8451-1059-4. £28.

Drugs Affecting Lipid Metabolism VIII. DAVID KRITCHESVY, WILLIAM L. HOLMES and RODOLFO PAOLETTI (eds). Plenum:1985. Pp.502. ISBN 0-306-41914-9. Np.

Environmental Carcinogens: Selected Methods of Analysis. Vol. 7: Some Volatile Halogenated Hydrocarbons. L. FISHBEIN and I.K. O'NEILL (eds). Oxford University Press:1985. Pp.479. Hbk ISBN 92-832-1168-5. £20.

Enzyme-Mediated Immunoassay. T.T. NGO and H.M. LENHOFF (eds). Plenum:1985. Pp.489. Hbk ISBN 0-306-42085-6. Np.

Experimental Assessments and Clinical Applications of Conditioned Food Aversions. NORMAN S. BRAVEMAN and PAUL BRONSTEIN (eds). New York Academy of Sciences:1985. Pp.441. Pbk ISBN 0-89766-281-4. Np.

Experiments in Plant Tissue Culture, 2nd Edn. By JOHN H. DODDS and LORIN W. ROBERTS. Cambridge University Press:1985. Pp.232. Hbk ISBN 0-521-30478-4; pbk ISBN 0-521-31516-6. Hbk £25, \$44.50; pbk £8.95, \$14.95.

Fish Aquaculture: Technology and Experiments. By CHRISTOPH MESKE. Pergamon:1985. Pp.237. Pbk ISBN 0-08-024191. Hbk £30, \$48; pbk £19, \$30.

Growth Factors and Transformation, Cancer Cells 3. J. FERAMISCO, B. OZANNE and C. STILES (eds). Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, New York 11724, USA:1985. Pp.450. Pbk ISBN 0-87969-178-6. \$70.

A Guide to Medical Endocrinology. By TAREK HASSAN. Macmillan:1985. Pp.269. Hbk ISBN 0-333-32027-1; pbk ISBN 0-333-32029-8. Hbk £17; pbk £8.95.

Human Prenatal Diagnosis. KAREN FILKINS and JOSEPH F. RUSSO (eds). Dekker:1985. Pp.424. Hbk ISBN 0-8247-7368-3. \$75 (US and Canada), \$90 (all other countries).

Human Sexuality: Psychosexual Effects of Disease. MARTIN FARBER (ed.). Collier Macmillan:1985. Pp.335. ISBN 0-02-336190-5. Hbk £35; pbk £25.

Infection, Immunity, and Blood Transfusion. R.Y. DODD and L.F. BARKER (eds). Alan R. Liss:1985. Pp.464. ISBN 0-8451-5032-4. £52.

The Inflammatory Reaction. By HENRY ZOLTAN MOVAT. Elsevier:1985. Pp.365. ISBN 0-444-80662-8. \$96.25.

Inherited Disorders of Vitamins and Cofactors. G.M. ADDISON, G. BARTLETT, R.A. HARKNESS and R.J. POLLITT (eds). MTP Press, Falcon House, Queen Sq., Lancaster, UK:1985. Pp.154. Hbk ISBN 0-85200-914-3. £29.95.

Integrated Mosquito Control Methodologies. Vol. 2. MARSHALL LAIRD and JAMES W. MILES (eds). Academic:1985. Pp.444. ISBN 0-12-434002-4. \$85, £65.50.

International Review of Experimental Pathology, Vol. 27. G.W. RICHTER and M.A. EPSTEIN (eds). Academic:1985. Pp.246. ISBN 0-12-364927-7. \$59, £59.

Introduction to Wildlife Management. By JAMES SHAW. McGraw-Hill:1985. Pp.316. ISBN 0-07-056481-7. £28.75.

In Vitro Fertilization and Embryo Transfer. MARKKU SEPPÄLÄ and R.G. EDWARDS (eds). New York Academy of Sciences:1985. Pp.619. Pbk ISBN 0-89766-279-4. Pbk np.

Issues and Reviews in Teratology, Vol. 3. HAROLD KALTER (ed.). Plenum:1985. Pp.299. Hbk ISBN 0-306-41988-2. Np.

Laser Application to Occlusive Vascular Disease. By MICHAEL W. BERNIS and MAHMOOD MIROSEINI (eds). Alan R. Liss:1985. Pp.344. ISBN 0-8451-0246-X. £22.

Ligand-Binder Assays: Labels and Analytical Strategies. By LARRY J. KRICKA. Dekker:1985. Pp.336. ISBN 0-8247-7420-5. \$65.

Macromolecules as Drugs and as Carriers for Biologically Active Materials. DAVID A. TIRRELL, L. GUY DONARUMA and ANNE B. TUREK (eds). New York Academy of Sciences:1985. Pp.458. Pbk ISBN 0-89766-287-3. Np.

Management of Arterial Hypertension: A Practical Guide for the Physician and Allied Health Workers. By F. GROSS, Z. PISA, T. STRASSER and A. ZANCHETTI. World Health Organization:1985. Pp.72. Pbk ISBN 92-4-154197-0. SwFr.11.

Memory Dysfunctions: An Integration of Animal and Human Research From Preclinical and Clinical Perspectives. DAVID S. OLTON, ELKANGAMZU and SUSANNE CORKIN (eds). New York Academy of Sciences:1985. Pp.553. Pbk ISBN 0-89766-283-0. Np.

Modern Approaches to the Dementias. Part I Etiology and Pathophysiology. Part II Clinical and Therapeutic Aspects. F. CLIFFORD ROSE (ed.). Karger:1985. Part I ISBN 3-8055-3980-0. Part II ISBN 3-8055-3981-9. Np.

No Magic Bullet: A Social History of Venereal Disease in the United States Since 1880. By ALLAN M. BRANDT. Oxford University Press:1985. Pp.245. ISBN 0-19-503469-4. £18.50.

Osteoporosis. 2 Vols. C. CHRISTIANSEN et al. (eds). Copenhagen International Symposium on Osteoporosis, Glostrup Hospital, University of Copenhagen, 2600 Glostrup, Denmark:1984. Vol. 1 Pp.536. Vol. 2 Pp.842. Np.

Pathophysiological Aspects of Cancer Epidemiology. G. MATHE and R. REIZENSTEIN (eds). Pergamon:1985. Pp.264. ISBN 0-08-030780-9. £50, \$81.50.

Patty's Industrial Hygiene and Toxicology, Vol. 3B 2nd Edn. Theory and Rationale of Industrial Hygiene Practice: Biological Responses. LEWIS J. CRALLEY and LESTER V. CRALLEY (eds). Wiley:1985. Pp.753. ISBN 0-471-82333-3. £92.

Pediatric Thyroidology. F. DELANGE et al. (eds). Karger:1985. Pp.412. ISBN 3-8055-3777-8. DM330, \$117.25.

Peptide Hormones as Mediators in Immunology and Oncology. R.D. HESCH (ed.). Raven Press:1985. Pp.254. Hbk ISBN 0-89004-609-3. \$46.50.

The Pharmacological Basis of Therapeutics. ALFRED GOODMAN GILMAN, LOUIS S. GOODMAN, THEODORE W. RALL and FERID MURAD (eds). Collier Macmillan, New York:1985. Pp.1839. Hbk ISBN 0-02344710-9. £65.

Psychophysiology of Cardiovascular Control: Models, Methods and Data. J.F. ORLEBEKE, G. MULDER and L.J.P. VAN DOORNEN (eds). Plenum:1985. Pp.959. ISBN 0-306-42007-4. Np.

Pulmonary Circulation in Chronic Lung Diseases: Experimental and Clinical Approach. J. WIDIMSKÝ, J. HERGET and J. MICZCH (eds). Karger:1985. Pp.192. ISBN 3-8055-3961-4. Np.

Reproductive Toxicology. ROBERT L. DIXON (ed.). Raven:1985. Pp.368. ISBN 0-89004-474-0. \$76.50.

Retinal Degeneration. Experimental and Clinical Studies. MATTHEW M. LAVAIL, JOE G. HOLLYFIELD and ROBERT E. ANDERSON (eds). Alan R. Liss:1985. Pp.485. Hbk ISBN 0-8451-4200-3. £53.

Soil Fertility and Fertilizers, 4th Edn. By SAMUEL L. TISDALE, WERNER L. NELSON and JAMES D. BEATON. Macmillan, NY:1985. Pp.754. Pbk ISBN 0-02-946760-8. Pbk £14.95.

Testicular Cancer and Other Tumors of the Genitourinary Tract. M. PAVONE-MACALUSO, P.H. SMITH and M.A. BAGSHAW (eds). Plenum:1985. Pp.537. ISBN 0-306-41906-8. Np.

Third European Congress on Biotechnology. Proceedings Vol. IV. EUROPEAN FEDERATION OF BIOTECHNOLOGY. VCH Publishers:1985. Pp.592. ISBN 0-89573-524-5. Np.

Toxicology of the Eye, Ear, and Other Special Senses. A. WALLACE HAYES (ed.). Raven Press:1985. Pp.254. Hbk ISBN 0-89004-840-1. Np.

Triangle. Sandoz Journal of Medical Science. Vol. 24, No. 1/2: Obstructive Lung Disease, Hypertension, Geriatric Endocrinology. Sandoz Pharmaceuticals:1985. Pp.70. Pbk np.

Understanding Dental Caries: Prevention. By GORDON NIKIFORUK. Karger:1985. Pp.288. ISBN 3-8055-3905-3. DM69, \$24.75.

Urolithiasis and Related Clinical Research. PAUL O. SCHWILLE et al. (eds). Plenum:1985. Pp.997. ISBN 0-306-41931-9. Np.